LASERS

THEORY AND APPLICATIONS

K. Thyagarajan
and
A.K. Ghatak
Indian Institute of Technology
New Delhi India

Preface

Since their invention in 1960, lasers have assumed tremendous importance in the fields of science and engineering because of their use both in basic research and in various technological applications The present book is an attempt at a coherent presentation of the basic physics behind the working of lasers along with some of their important applications

A major portion of the book evolved from lectures given by the authors to senior undergraduate and graduate students at the Indian Institute of Technology, New Delhi A part of the book has also been used in a few summer/winter schools organized by the Institute

In the first part of the book, an attempt is made to present, in a coherent fashion, the basic theory behind laser operation The topics are introduced from first principles so that the book can be used for self-study We do not discuss the maser principle, however, the Nobel lecture of Townes gives a nice discussion of this and is reproduced in Part III of this book The fully quantum mechanical theory is also not given as it is beyond the scope of the present book

In the second part of the book, we discuss, in reasonable detail, some of the important applications of lasers, including topics such as spatial frequency filtering, holography, laser-induced fusion, light wave communications, and applications of lasers in pure sciences and in industry Although there are many more applications which are not included in this book, we feel that we have covered the most important applications It was not possible to go into the details of the various applications because for each of the applications separate books could be (and have been) written

We feel that the reader should have some sense of perspective of the history of the development of the laser One obvious way would be to introduce the reader to some of the original papers, unfortunately these papers are usually not easy to read and involve considerable mathematical complexity We felt that the Nobel lectures of Townes, Basov, and

Prochorov† would convey the development of the subject in a manner that could not possibly be matched, and therefore, in the third part of the book, we reproduce these Nobel lectures We also reproduce the Nobel lecture of Gabor‡ because in it Gabor has beautifully distilled the physics and the important applications of holography

We hope that the book will be of use to scientists and engineers who plan to study or teach the basic physics behind the operation of lasers along with their important applications

We will be most grateful for comments from readers

New Delhi

K Thyagarajan
A K Ghatak

† Townes Basov and Prochorov were jointly awarded the 1964 Nobel Prize for Physics for their fundamental work relating to the maser–laser principle

‡ Dennis Gabor was awarded the 1971 Nobel Prize for Physics for the invention of holography

Acknowledgments

We are greatly indebted to Dr S V Lawande of Bhabha Atomic Research Center, Bombay, for writing Section 13 9 on laser isotope separation We are also grateful to Professor P K Kaw and Dr Y Satya for their comments on Chapter 11

We are indebted to the various authors and publishers for their permission to use various figures appearing in the book and to Elsevier Publishing Company for permitting us to reproduce the Nobel lectures We are very grateful to Dr A G Chynoweth, Professor Gurbax Singh, Dr H Kogelink, Dr T A Leonard, Dr D F Nelson, Dr R A Phillips, Dr R W Terhune, Dr L A Weaver, Ferranti Ltd , and the United States Information Service, New Delhi, for providing glossy prints of some of the photographs appearing in the book

We are grateful to Professor M S Sodha for his interest and encouragement in this endeavor We are also grateful to Drs Anurag Sharma, Aruna Rohra, Arun Kumar, R D Bahuguna, Enakshi Sharma, I C Goyal, B D Gupta, K K Gupta, Iqbal Hussain, Kehar Singh, and B P Pal for their help at various stages during the preparation of the manuscript and for many stimulating discussions One of us (A K G) is also grateful to Professor Peter Herczfeld of Drexel University for the hospitality extended to him at Drexel, where a part of the manuscript was written

We are grateful to Mr S D Malik, Mr T N Gupta, and Mr N S Gupta for their help in the preparation of the manuscript

Finally, we would like to record our gratitude to our Director, Professor O P Jain, for his interest and support of this work

Contents

1

Introduction

An atomic system is characterized by discrete energy states, and usually the atoms exist in the lowest energy state, which is normally referred to as the ground state An atom in a lower energy state may be excited to a higher energy state through a variety of processes One of the important processes of excitation is through collisions with other particles The excitation can also occur through the absorption of electromagnetic radiation of proper frequencies such a process is known as stimulated absorption or simply as absorption On the other hand, when the atom is in the excited state, it can make a transition to a lower energy state through the emission of electromagnetic radiation, however, in contrast to the absorption process, the emission process can occur in two different ways

(i) The first is referred to as spontaneous emission in which an atom in the excited state emits radiation even in the absence of any incident radiation It is thus not stimulated by any incident signal but occurs spontaneously Further the rate of spontaneous emissions is proportional to the number of atoms in the excited state [see Eq (3 2-3)]

(ii) The second is referred to as stimulated emission, in which an incident signal of appropriate frequency triggers an atom in an excited state to emit radiation The rate of stimulated emission (or absorption) depends both on the intensity of the external field and also on the number of atoms in the upper state The net stimulated transition (stimulated absorption and stimulated emission) depends on the difference in the number of atoms in the excited and the lower states, unlike the case of spontaneous emission which depends only on the population of the excited state

The fact that there should be two kinds of emissions—namely, spontaneous and stimulated—was first predicted by Einstein in 1917 The consideration which led to this prediction was the description of thermodynamic equilibrium between atoms and the radiation field Einstein

(1917) showed that both spontaneous and stimulated emissions are necessary to obtain Planck's radiation law, this is discussed in Section 3 2 The quantum mechanical theory of spontaneous and stimulated emission is discussed in Section 8 5

The phenomenon of stimulated emission was first used by Townes in 1954 in the construction of a microwave amplifier device called the maser†, which is an acronym for microwave amplification by stimulated emission of radiation At about the same time a similar device was also proposed by Prochorov and Basov The maser principle was later extended to the optical frequencies by Schawlow and Townes in 1958, which led to the device now known as the laser In fact "laser" is an acronym for light amplification by stimulated emission of radiation The first successful operation of a laser device was demonstrated by Maiman in 1960 using ruby crystal (see Section 9 3) Within a few months of operation of the device, Javan and his associates constructed the first gas laser, namely, the He–Ne laser (see Section 9 4) Since then, laser action has been obtained in a large variety of materials including liquids, ionized gases, dyes, semiconductors etc (see Chapter 9)

The three main components of any laser device are the active medium, the pumping source, and the optical resonator The active medium consists of a collection of atoms, molecules, or ions (in solid, liquid, or gaseous form), which acts as an amplifier for light waves For amplification, the medium has to be kept in a state of population inversion, i e , in a state in which the number of atoms in the upper energy level is greater than the number of atoms in the lower energy level The pumping mechanism provides for obtaining such a state of population inversion between a pair of energy levels of the atomic system When the active medium is placed inside an optical resonator, the system acts as an oscillator

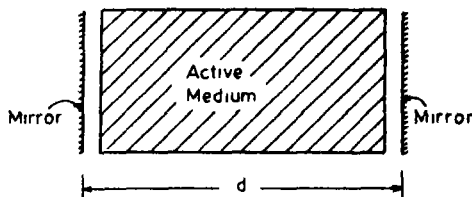
After developing the necessary quantum mechanics in Chapter 2, in Chapter 3 we give the original argument of Einstein regarding the presence of both spontaneous and stimulated emissions and obtain expressions for the rate of absorption and emission using a semiclassical theory We also consider the interaction of an atom with electromagnetic radiation over a band of frequencies and obtain the gain (or loss) coefficient as the beam propagates through the active medium

Under normal circumstances, there is always a larger number of atoms in the lower energy state as compared to the excited energy state and an electromagnetic wave passing through such a collection of atoms

† A nice account of the maser device is given in the Nobel lecture of Townes, which is reproduced in Part III of this book

would get attenuated rather than amplified. Thus, in order to have amplification, one must have population inversion In Chapter 4, we discuss the three-level and four-level systems for obtaining population inversion between two states of the system It is shown that in order to obtain a population inversion, the transition rates of the various levels must satisfy certain conditions We also obtain the pumping powers required for obtaining population inversion in three- and four-level systems and show that it is in general much easier to obtain inversion in a four-level system as compared to a three-level system In Chapter 5 we give the semiclassical theory of laser operation and show that the amplification process due to stimulated transitions is phase coherent—i e., an electromagnetic wave passing through an inverted medium gets amplified and the phase of the wave is changed by a constant amount, the gain depends on the amount of inversion

A medium with population inversion is capable of amplification, but if the medium is to act as an oscillator, a part of the output energy must be fed back into the system† Such a feedback is brought about by placing the active medium between a pair of mirrors which are facing each other (see Fig 1 1) The mirrors could be either plane or curved Such a system formed by a pair of mirrors is referred to as a resonator The sides of the cavity are, in general, open and hence such resonators are also referred to as open resonators In Chapter 6 we give a detailed account of optical resonators and obtain the oscillation frequencies of the modes of the resonator The different field patterns of the various modes are also obtained

Fig. 1 1 A plane parallel resonator consisting of a pair of plane mirrors facing each other The active medium is placed inside the cavity One of the mirrors is made partially transmitting to couple out the laser beam

Because of the open nature of the resonators, all modes of the resonator are lossy due to the diffraction spillover of energy from the mirrors In addition to this basic loss, the scattering in the laser medium, the absorption at the mirrors, and the loss due to output coupling of the

† Since some of the energy is coupled back to the system, it is said to act as an oscillator Indeed in the early stages of the development of the laser, there was a move to change its name to loser which is an acronym for light oscillation by stimulated emission of radiation Since it would have been difficult to obtain research grants on losers, it was decided to retain the name laser

mirrors also lead to losses In an actual laser, the modes that keep oscillating are those for which the gain provided by the laser medium compensates for the losses When the laser is oscillating in steady state, the losses are exactly compensated by the gain Since the gain provided by the medium depends on the amount of population inversion, there is a critical value of population inversion beyond which the particular mode would oscillate in the laser If the population inversion is less than this value, the mode cannot oscillate The critical value of population inversion is also called the threshold population inversion In Chapters 3 and 5 we obtain explicit expressions for the threshold population inversion in terms of the parameters of the laser medium and the resonator

The quantum mechanical theory of spontaneous and stimulated emission is discussed in Chapter 8, the necessary quantum mechanics is given in Chapter 7

The onset of oscillations in a laser cavity can be understood as follows Through some pumping mechanism one creates a state of population inversion in the laser medium placed inside the resonator system Thus the medium is prepared to be in a state in which it is capable of coherent amplification over a specified band of frequencies The spontaneous emission occurring inside the resonator cavity excites the various modes of the cavity For a given population inversion, each mode is characterized by a certain amplification coefficient due to the gain and a certain attenuation coefficient due to the losses in the cavity The modes for which the losses in the cavity exceed the gain will die out On the other hand, the modes whose gain is higher than the losses get amplified by drawing energy from the laser medium The amplitude of the mode keeps on increasing till nonlinear saturation depletes the upper level population to a value when the gain equals the losses and the mode oscillates in steady state In Chapter 4 we study the change in the energy in a mode as a function of the rate of pumping and have shown that as the pumping rate passes through the threshold value, the energy contained in a mode rises very steeply and the steady state energy in a mode above threshold is orders of magnitude greater than the energy in the same mode below threshold Since the laser medium provides gain over a band of frequencies, it may happen that many modes have a gain higher than the loss, and in such a case the laser oscillates in a multimode fashion In Chapter 6 we also briefly discuss various techniques for selecting a single mode oscillation of the cavity

The light emitted by ordinary sources of light, like the incandescent lamp, is spread over all directions and is usually over a large range of wavelengths In contrast, the light from a laser could be highly monochromatic and highly directional Because of the presence of the optical

cavity, only certain discrete frequencies can oscillate in the cavity In addition, when the laser is oscillating in steady state the losses are exactly compensated by the gain provided by the medium and the wave coming out of the laser can be represented as a nearly continuous wave The ultimate monochromaticity is determined by the spontaneous emissions occurring inside the cavity because the radiation coming out of the spontaneous emissions is incoherent The notion of coherence is discussed in Chapter 9 and the expression for the ultimate monochromaticity of the emitted radiation is discussed in Chapter 6 In practice, the monochromaticity is limited due to external factors like temperature fluctuations, mechanical oscillations of the optical cavity, etc The light coming out of the laser which is oscillating in a single mode is also composed of a well-defined wave front This comes about because of the effects of propagation and diffraction inside the resonator cavity This property is also discussed in greater detail in Section 9 2

In Chapter 9 we briefly discuss some of the different kinds of lasers and the special properties possessed by laser light and the concept of temporal and spatial coherence

In Chapters 10–14 we discuss some of the important applications of lasers which have come about because of the special properties of lasers Finally in Part III of the book we reprint the Nobel Lectures of Townes, Prochorov, Basov, and Gabor Townes, Prochorov, and Basov were awarded the 1964 Nobel Prize for physics for their invention of the laser devices The Nobel lectures of Townes and Prochorov discuss the basic principles of the maser and the laser whereas the Nobel lecture of Basov gives a detailed account of semiconductor lasers Probably the most important and interesting application of the laser is holography, which has been beautifully discussed in the Nobel lecture by Dennis Gabor, this is also reproduced in Part III of the book

PART I

Basic Laser Theory

2

Elements of Quantum Mechanics

2.1. Introduction

In this chapter we discuss very briefly the basic principles of quantum mechanics which are used in later chapters At places, the chapter will appear disconnected, this is inevitable because the subject of quantum mechanics is so vast that it is impossible to present the basic concepts in a coherent fashion in one tiny chapter! Nevertheless, whatever we discuss we will try to do from first principles

We first consider the propagation of a wave packet, which is followed by the Schrodinger equation, the physical interpretation of the wave function, and the uncertainty principle As an illustration of how the Schrodinger equation is solved, we discuss the one-dimensional harmonic oscillator problem in detail, and for the hydrogen atom problem, we just present the results Several other "solvable" problems are briefly discussed at the end of the chapter

We may mention here that in Chapter 7 we discuss the Dirac bra and ket algebra, which is another formulation of quantum mechanics

2 2 Wave Packet

There are many experimental results which show that atomic objects (like electrons, protons, neutrons, α particles, etc) exhibit both wave and particle properties Indeed the de Broglie relation

$$\lambda = \frac{h}{p} \qquad (2\,2\text{-}1)$$

between wavelength λ and momentum p suggests that a particle whose position is relatively sharply defined can be represented by a wave

packet†, in Eq (2 2-1) h ($\approx 6.627 \times 10^{-34}$ J sec) represents Planck's constant Thus the behavior of the particle should be describable by a wave function $\Psi(\mathbf{r}, t)$ whose magnitude is large in regions where the probability of occurrence of the particle is large, in other regions where the particle is less likely to be found, the magnitude of Ψ is small Before we construct the wave packet, we note that the simplest type of wave is a plane wave

$$\exp[i(\mathbf{k} \cdot \mathbf{r} - \omega t)] \tag{2 2-2}$$

which represents a disturbance of wavelength $\lambda = 2\pi/k$ traveling in the direction of its wave vector $\mathbf{k}$ with angular frequency ω Now, as established by Einstein's explanation of the photoelectric effect, the energy E of the particle is related to the frequency ω by the following equation

$$E = \hbar\omega \tag{2 2-3}$$

where $\hbar \equiv h/2\pi$ Further, from the de Broglie relation [Eq (2 2-1)], the momentum $\mathbf{p}$ and wave number $\mathbf{k}$ are related by the following equation

$$\mathbf{p} = \hbar\mathbf{k} \tag{2 2-4}$$

Restricting our discussion to one dimension, we construct a wave packet by superposition of plane waves

$$\Psi(x, t) = \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{+\infty} A(k) \exp[i(kx - \omega t)]\, dk \tag{2 2-5}$$

The quantity $A(k)$ is determined from the initial form of the wave packet

$$\Psi(x, 0) = \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{+\infty} A(k) \exp(ikx)\, dk \tag{2 2-6}$$

Thus, by taking the inverse Fourier transform (see Appendix A)

$$A(k) = \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{+\infty} \Psi(x, 0) \exp(-ikx)\, dx \tag{2 2-7}$$

Using Eqs (2 2-3) and (2 2-4), Eq (2 2-5) may be written in the form

$$\Psi(x, t) = \frac{1}{(2\pi\hbar)^{1/2}} \int_{-\infty}^{+\infty} a(p) \exp\left[\frac{i}{\hbar}(px - Et)\right] dp \tag{2 2-8}$$

† A wave packet is a type of wave motion in which the amplitude of the wave is large in a small region and negligibly small in the rest of space mathematically, it can be obtained by superposition of plane waves

where the factor $(2\pi\hbar)^{-1/2}$ outside the integral has been introduced such that $a(p)$ is given by the symmetrical relation

$$a(p) = \frac{1}{(2\pi\hbar)^{1/2}} \int_{-\infty}^{+\infty} \Psi(x, 0) \exp\left(-\frac{i}{\hbar} px\right) dx \qquad (2.2\text{-}9)$$

It is interesting to mention that $a(p)$ is also given by the following equation

$$a(p) = \frac{1}{(2\pi\hbar)^{1/2}} \int_{-\infty}^{+\infty} \Psi(x, t) \exp\left[-\frac{i}{\hbar} (px - Et)\right] dx \qquad (2.2\text{-}10)$$

For a free particle, the energy and momentum are related by the equation

$$E = \frac{p^2}{2m} \qquad (2.2\text{-}11)$$

Example As an example†, we consider the propagation of a one-dimensional Gaussian wave packet‡ for which

$$\Psi(x, t = 0) = \frac{1}{(\pi\sigma^2)^{1/4}} \exp\left[-\frac{(x - x_0)^2}{2\sigma^2}\right] \exp\left(\frac{i}{\hbar} p_0 x\right) \qquad (2.2\text{-}12)$$

which represents a particle localized within $\Delta x \sim \sigma$ and moving with an average momentum§ p_0. The factor $(\pi\sigma^2)^{1/4}$ is such that

$$\int_{-\infty}^{+\infty} |\Psi|^2 \, dx = 1 \qquad (2.2\text{-}13)$$

We will show later that $|\Psi|^2 \, dx$ can be interpreted as the probability of finding the particle within x and $x + dx$ (see Section 2.4), thus Eq. (2.2-13) is simply the normalization condition implying that the particle has to be found somewhere. Using Eq. (2.2-9) and carrying out straightforward integration one obtains (see Appendix B)

$$a(p) = \left(\frac{\sigma^2}{\pi\hbar^2}\right)^{1/4} \exp\left[-\frac{\sigma^2}{2\hbar^2}(p - p_0)^2 - \frac{i}{\hbar}(p - p_0)x_0\right] \qquad (2.2\text{-}14)$$

Physically, as we shall show later (see second footnote on p. 18), $|a(p)|^2 \, dp$ represents the probability of finding the momentum (of the particle) between p and $p + dp$, thus, Eq. (2.2-14) implies that the uncertainty in momentum Δp is $\sim\hbar/\sigma$ (implying $\Delta x \, \Delta p \sim \hbar$—consistent with the uncertainty principle, see Section 2.7). If we now substitute for $a(p)$ in Eq. (2.2-8) and carry out the integration, we would get an analytical expression for $\Psi(x, t)$ (see Appendix B), the real part of which is plotted in Fig. 2.1. The figure clearly demonstrates the propagation of a wave packet with time. It also shows the broadening of the packet, this is a manifestation of the fact that the medium is dispersive, which follows

† This example may be read after going through the entire chapter.

‡ The details have been considered in Appendix B.

§ The average momentum is defined through Eq. (2.5-8).

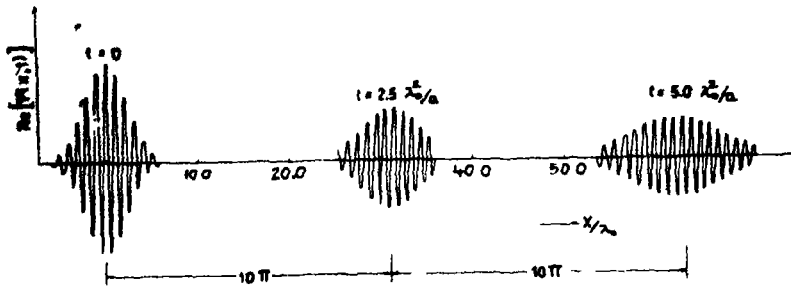



Fig 2 1 The broadening of a Gaussian pulse in a dispersive medium characterized by Eq (2 2-16) Numerical calculations courtesy of Ms K K Sankari

from the relation

$$\hbar\omega = E = \frac{p^2}{2m} = \frac{\hbar^2 k^2}{2m} \tag{2 2-15}$$

Thus

$$\omega = \frac{\hbar}{2m} k^2 \tag{2 2-16}$$

implying that the ω–k relationship is not linear which makes the medium dispersive

The group velocity is given by

$$v_g = \frac{d\omega}{dk} = \frac{\hbar k}{m} = \frac{p}{m} \tag{2 2-17}$$

and if $a(p)$ is fairly sharply peaked around $p = p_0$ then the packet moves with an average group velocity given by

$$v_g \approx \frac{p_0}{m} \tag{2 2-18}$$

as can be seen from the figure

The three-dimensional generalization of Eqs (2 2-8) and (2 2-9) would be

$$\Psi(\mathbf{r}, t) = \frac{1}{(2\pi\hbar)^{3/2}} \iiint a(\mathbf{p}) \exp\left[\frac{i}{\hbar}(\mathbf{p} \cdot \mathbf{r} - Et)\right] d\mathbf{p} \tag{2 2-19}$$

and

$$a(\mathbf{p}) = \frac{1}{(2\pi\hbar)^{3/2}} \iiint \Psi(\mathbf{r}, 0) \exp\left(-\frac{i}{\hbar}\mathbf{p} \cdot \mathbf{r}\right) d\mathbf{r} \tag{2 2-20}$$

2.3. The Schrödinger Equation

If we successively differentiate Eq (2 2-8), we obtain the relations

$$i\hbar \frac{\partial \Psi}{\partial t} = \frac{1}{(2\pi\hbar)^{1/2}} \int_{-\infty}^{+\infty} Ea(p) \exp\left[\frac{i}{\hbar}(px - Et)\right] dp \tag{2 3-1}$$

$$-i\hbar \frac{\partial \Psi}{\partial x} = \frac{1}{(2\pi\hbar)^{1/2}} \int_{-\infty}^{+\infty} pa(p) \exp\left[\frac{i}{\hbar}(px - Et)\right] dp \tag{2 3-2}$$

$$-\hbar^2 \frac{\partial^2 \Psi}{\partial x^2} = \frac{1}{(2\pi\hbar)^{1/2}} \int_{-\infty}^{\infty} p^2 a(p) \exp\left[\frac{i}{\hbar}(px - Et)\right] dp \tag{2 3-3}$$

Using Eq (2 2-11) we obtain

$$i\hbar \frac{\partial \Psi}{\partial t} = -\hbar^2 \frac{\partial^2 \Psi}{\partial x^2} \tag{2 3-4}$$

which is known as the one-dimensional Schrodinger equation for a free particle Further, for a plane wave

$$\Psi = \exp\left[\frac{i}{\hbar}(px - Et)\right] \tag{2 3-5}$$

and

$$i\hbar \frac{\partial \Psi}{\partial t} = E\Psi \tag{2 3-6}$$

$$-i\hbar \frac{\partial \Psi}{\partial x} = p\Psi \tag{2 3-7}$$

which suggests that, at least for a free particle, the energy and momentum can be represented by differential operators given by

$$E \to i\hbar \frac{\partial}{\partial t}, \qquad p \to -i\hbar \frac{\partial}{\partial x} \tag{2 3-8}$$

If we use the differential operator representation, Eq (2 3-4) can be written in the form

$$E\Psi = \frac{p^2}{2m}\Psi \tag{2 3-9}$$

We next consider the particle to be in a force field characterized by the potential energy $V(x)$, thus, classically, the total energy of the system is given by

$$E = \frac{p^2}{2m} + V(x) \tag{2 3-10}$$

If we now assume p and E to be represented by the differential operators, the equation

$$E\Psi = \left[\frac{p^2}{2m} + V(x)\right]\Psi \tag{2 3-11}$$

would assume the form

$$i\hbar\frac{\partial\Psi}{\partial t} = \left[-\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2} + V(x)\right]\Psi \tag{2 3-12}$$

which represents the one-dimensional Schrodinger equation The three-dimensional generalization is quite straightforward, instead of Eq (2 3-8), we have

$$E \rightarrow i\hbar\frac{\partial}{\partial t} \tag{2 3-13}$$

$$p_x \rightarrow -i\hbar\frac{\partial}{\partial x}, \qquad p_y \rightarrow -i\hbar\frac{\partial}{\partial y}, \qquad p_z \rightarrow -i\hbar\frac{\partial}{\partial z} \tag{2 3-14}$$

Thus the equation [cf Eq (2 3-11)]

$$E\Psi = \left[\frac{1}{2m}(p_x^2 + p_y^2 + p_z^2) + V(\mathbf{r})\right]\Psi \tag{2 3-15}$$

assumes the form

$$i\hbar\frac{\partial\Psi}{\partial t} = H\Psi \tag{2 3-16}$$

where

$$H = \frac{p^2}{2m} + V(\mathbf{r}) = -\frac{\hbar^2}{2m}\nabla^2 + V(\mathbf{r}) \tag{2 3-17}$$

is an operator and represents the Hamiltonian of the system Equations (2 3-16) and (2 3-17) represent the three-dimensional Schrodinger equation

When the Hamiltonian, H, is independent of time†, Eq (2 3-16) can be solved by using the method of separation of variables

$$\Psi(\mathbf{r}, t) = \psi(\mathbf{r})T(t) \tag{2 3-18}$$

† Whenever we are considering bound states of a system (like those of the hydrogen atom or that of the harmonic oscillator) the Hamiltonian is independent of time however, for problems like when we consider the interaction of an atom with the radiation field the Hamiltonian is not independent of time (see e g Section 3 3)

Substituting in Eq (2 3-16) and dividing by Ψ, we obtain

$$i\hbar \frac{1}{T(t)} \frac{dT}{dt} = \frac{1}{\psi}\left[-\frac{\hbar^2}{2m}\nabla^2\psi + V(\mathbf{r})\psi\right] = E \tag{2 3-19}$$

where E is a constant (and now a number) Thus

$$\frac{dT}{dt} + \frac{i}{\hbar} ET(t) = 0 \tag{2 3-20}$$

giving

$$T(t) \sim \exp\left(-\frac{i}{\hbar} Et\right) \tag{2 3-21}$$

Further, Eq (2 3-19) gives us

$$-\frac{\hbar^2}{2m}\nabla^2\psi + V\psi = E\psi \tag{2 3-22}$$

or,

$$H\psi = E\psi \tag{2 3-23}$$

which is essentially an eigenvalue equation For ψ to be "well behaved,"† the quantity E takes some particular values (see, e g , the harmonic oscillator problem—Section 2 8), these are known as the energy eigenvalues and the corresponding forms of ψ are known as eigenfunctions The solution

$$\Psi_n(\mathbf{r}, t) = \psi_n(\mathbf{r}) \exp(-iE_n t/\hbar) \tag{2 3-24}$$

is said to describe a stationary state, here the subscript n refers to a particular eigenvalue E_n Thus

$$H\psi_n = E_n\psi_n \tag{2 3-25}$$

The eigenfunctions $\psi_n(\mathbf{r})$ from a complete set, i e , an arbitrary well-behaved function Φ, can be expanded in terms of the functions $\psi_n(\mathbf{r})$

$$\Phi(\mathbf{r}) = \sum_n a_n\psi_n(\mathbf{r}) \tag{2 3-26}$$

where $\sum$ represents a summation over the discrete states and integration over the continuum states For example, for the harmonic oscillator

† By "well-behaved" we imply functions which are single valued, square-integrable (i e , $\int |\Phi|^2\, d\tau$ should exist) etc

problem only discrete states exist, whereas for the hydrogen atom problem discrete and continuum states exist

2.4. Physical Interpretation of Ψ and its Normalization

We rewrite the Schrodinger equation

$$i\hbar\frac{\partial\Psi(\mathbf{r},t)}{\partial t} = -\frac{\hbar^2}{2m}\nabla^2\Psi + V(\mathbf{r})\Psi \tag{2 4-1}$$

along with its complex conjugate

$$-i\hbar\frac{\partial\Psi^*(\mathbf{r},t)}{\partial t} = -\frac{\hbar^2}{2m}\nabla^2\Psi^* + V(\mathbf{r})\Psi^* \tag{2 4-2}$$

If we multiply Eq (2 4-1) by Ψ^* and Eq (2 4-2) by Ψ and subtract, we obtain

$$i\hbar\left(\Psi^*\frac{\partial\Psi}{\partial t} + \Psi\frac{\partial\Psi^*}{\partial t}\right) = -\frac{\hbar^2}{2m}(\Psi^*\nabla^2\Psi - \Psi\nabla^2\Psi^*) \tag{2 4-3}$$

Remembering that $\nabla^2 = \partial^2/\partial x^2 + \partial^2/\partial y^2 + \partial^2/\partial z^2$, we may rewrite the above equation in the form

$$\frac{\partial}{\partial t}(\Psi^*\Psi) + \frac{i\hbar}{2m}\left[\frac{\partial}{\partial x}\left(\Psi\frac{\partial\Psi^*}{\partial x} - \Psi^*\frac{\partial\Psi}{\partial x}\right) + \frac{\partial}{\partial y}\left(\Psi\frac{\partial\Psi^*}{\partial y} - \Psi^*\frac{\partial\Psi}{\partial y}\right) + \frac{\partial}{\partial z}\left(\Psi\frac{\partial\Psi^*}{\partial z} - \Psi^*\frac{\partial\Psi}{\partial z}\right)\right] = 0$$

or

$$\frac{\partial\rho}{\partial t} + \nabla\cdot\mathbf{J} = 0 \tag{2 4-4}$$

where

$$\rho = \Psi^*\Psi \tag{2 4-5}$$

$$\nabla\cdot\mathbf{J} = \frac{\partial J_x}{\partial x} + \frac{\partial J_y}{\partial y} + \frac{\partial J_z}{\partial z} \tag{2 4-6}$$

$$J_x = \frac{i\hbar}{2m}\left(\Psi\frac{\partial\Psi^*}{\partial x} - \Psi^*\frac{\partial\Psi}{\partial x}\right) \tag{2 4-7}$$

and similar expressions for J_y and J_z Equation (2 4-4) is the equation of

continuity in fluid dynamics and can be physically interpreted by considering a moving gas with ρ representing the number of particles per unit volume and $\mathbf{J}$ representing the current density [see, e g , Tait (1964)] Thus, if we normalize Ψ such that†

$$\iiint_{-\infty}^{+\infty} \Psi^*\Psi \, d\tau = 1 \qquad (2\,4\text{-}8)$$

we may associate

$$\rho = \Psi^*\Psi \qquad (2\,4\text{-}9)$$

with *position probability density* and $\mathbf{J}$ with *probability current density* This implies that $\Psi^*\Psi\, d\tau$ represents the probability of finding the particle in the volume element $d\tau$ Further, for an infinitely extended plane wave

$$\Psi = \exp\left[\frac{i}{\hbar}(\mathbf{p}\cdot\mathbf{r} - Et)\right] \qquad (2\,4\text{-}10)$$

the current density $\mathbf{J}$ can be easily calculated to give

$$\mathbf{J} = \frac{\mathbf{p}}{m} = \mathbf{v} \qquad (2\,4\text{-}11)$$

which is just the current for a beam of particles of unit density‡ ($\Psi^*\Psi = 1$) and velocity $\mathbf{v}$

2 5 Expectation Values of Dynamical Quantities

The interpretation of Ψ in terms of the position probability density allows us to calculate the expectation value of measurable quantities For example, the expectation value of the x coordinate is given by

$$\langle x \rangle = \frac{\iiint x\Psi^*\Psi\, d\tau}{\iiint \Psi^*\Psi\, d\tau} = \iiint \Psi^*(\mathbf{r}, t)x\Psi(\mathbf{r}, t)\, d\tau \qquad (2\,5\text{-}1)$$

where the integration is over the entire space and in the last step we have

† For all states for which $\int \Psi^*\Psi\, d\tau$ exists this normalization is always possible because if Ψ is a solution of Eq (2 4-1) then any multiple of Ψ is also a solution and we may always choose the multiplicative constant such that Eq (2 4-8) is satisfied

‡ It may be noted that the plane wave is not normalizable this is due to the fact that an infinitely extended plane wave corresponds to a constant probability density everywhere

assumed the wave function to be normalized Similarly, we may write for $\langle y\rangle$ and $\langle z\rangle$ and also for the expectation value of the potential energy V

$$\langle V\rangle = \iiint \Psi^*(\mathbf{r},t)V(\mathbf{r})\Psi(\mathbf{r},t)\,d\tau \tag{2 5-2}$$

In order to obtain an expression for the expectation values of quantities like energy and momentum, we multiply the Schrodinger equation [Eq (2 4-1)] by Ψ^* and integrate it to obtain†

$$\int \Psi^* i\hbar\frac{\partial\Psi}{\partial t}\,d\tau = \int \Psi^*\left(-\frac{\hbar^2}{2m}\nabla^2\right)\Psi\,d\tau + \int \Psi^* V\Psi\,d\tau \tag{2 5-3}$$

The last term is simply $\langle V\rangle$, further, if we assume the reasonable requirement

$$\langle E\rangle = \left\langle\frac{p^2}{2m}\right\rangle + \langle V\rangle \tag{2 5-4}$$

then we may write

$$\langle E\rangle = \int \Psi^*\left(i\hbar\frac{\partial\Psi}{\partial t}\right)d\tau \tag{2 5-5}$$

$$\langle p_x^2\rangle = \int \Psi^*\left(-\hbar^2\frac{\partial^2\Psi}{\partial x^2}\right)d\tau \tag{2 5-6}$$

and similar expressions for $\langle p_y^2\rangle$ and $\langle p_z^2\rangle$ Equations (2 5-5) and (2 5-6) suggest that the expectation value of any dynamical quantity $\mathcal{O}$ is obtained by operating it on Ψ, premultiplying it by Ψ^*, and then integrating

$$\langle\mathcal{O}\rangle = \int \Psi^*\mathcal{O}\Psi\,d\tau \tag{2 5-7}$$

In particular‡

$$\langle p_x\rangle = \int \Psi^*\left(-i\hbar\frac{\partial\Psi}{\partial x}\right)d\tau \tag{2 5-8}$$

† From now on the single integral sign will be assumed to represent the three-dimensional integral over the entire space

‡ For the one-dimensional wave packet [see Eq (2 2 8)] it is easy to show that

$$\langle p\rangle = \int a^*(p)pa(p)\,dp$$

and

$$\langle p^2\rangle = \int a^*(p)p^2a(p)\,dp$$

Thus $|a(p)|^2\,dp$ can be interpreted as the probability for the momentum to be between p and $p + dp$

2.6 The Commutator

The commutator of two operators α and β is defined by the following equation

$$[\alpha, \beta] = \alpha\beta - \beta\alpha = -[\beta\ \alpha] \tag{2 6-1}$$

Now, the commutator of x and p_x operating on an arbitrary function Ψ is given by

$$[x\ p_x]\Psi = (xp_x - p_x x)\Psi = -i\hbar\left[x\frac{\partial\Psi}{\partial x} - \frac{\partial}{\partial x}(x\Psi)\right]$$
$$= i\hbar\Psi \tag{2 6-2}$$

Since Ψ is arbitrary, we obtain

$$[x\ p_x] = xp_x - p_x x = i\hbar \tag{2 6-3}$$

Similarly

$$[y, p_y] = [z, p_z] = i\hbar \tag{2 6-4}$$

However

$$[x, p_y] = [y, p_z] = \qquad = 0 \tag{2 6-5}$$

$$[x, y] = [y, z] = \qquad = 0 \tag{2 6-6}$$

and

$$[p_x, p_y] = [p_y, p_z] = \qquad = 0 \tag{2 6-7}$$

2 7 The Uncertainty Principle

One of the most important consequences of the wave–particle duality is the famous Heisenberg uncertainty principle In a sense, the uncertainty principle protects the wave–particle duality According to this principle, it is impossible to measure simultaneously a Cartesian component of the position and the corresponding momentum with arbitrary accuracy Indeed if Δx and Δp_x represent the uncertainties in the measurement of the x coordinate of the position and of the x component of the momentum, then

$$\Delta x\, \Delta p_x \geqslant h \tag{2 7-1}$$

Similarly

$$\Delta y\, \Delta p_y \geqslant h \quad \text{and} \quad \Delta z\, \Delta p_z \geqslant h$$

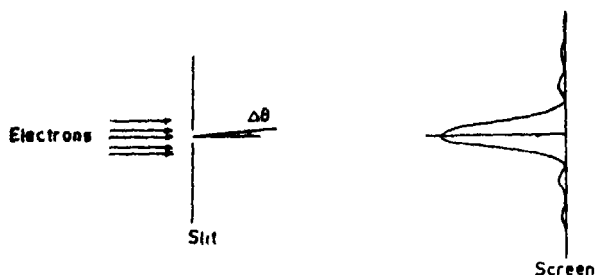



Fig 2 2 Diffraction of an electron beam by a long narrow slit of width d

In order to understand the inherent uncertainties in the simultaneous measurements of x and p_x, we consider the passage of electrons (from a very distant source) through a long narrow slit of width d (Fig 2 2) If the source (which emits the electrons) is sufficiently far away, the electron will have almost zero x component of the momentum *before* it entered the slit, i e , $\Delta p_x \approx 0$ We now try to limit the x coordinate of the electron by passing it through the slit thus

$$\Delta x \sim d \tag{2 7-2}$$

However, as soon as the electron passes through the slit, diffraction occurs, indeed the diffraction spreading (in the x direction) is approximately given by

$$\Delta\theta \sim \frac{\lambda}{d} \tag{2 7-3}$$

Thus the electron acquires a momentum† in the x direction which is given by

$$\Delta p_x \sim p_x \sim p\,\Delta\theta \sim p\frac{\lambda}{d} \tag{2 7-4}$$

† This can be also seen by assuming the wave packet (associated with the electron) to be approximately given by

$$\Psi(x\ 0) = \begin{cases} 1/d^{1/2} & |x| < d/2 \\ 0 & |x| > d/2 \end{cases}$$

Thus

$$\int |\Psi(x\ 0)|^2\, dx = 1$$

and

$$a(p_x) = (2\pi\hbar)^{-1/2} \int \Psi(x\ 0) \exp\left(-\frac{i}{\hbar} p_x x\right) dx$$

$$= (d/2\pi\hbar)^{1/2} \frac{\sin(p_x d/2\hbar)}{(p_x d/2\hbar)}$$

The distribution $|a(p_x)|^2\, dp_x$ gives the intensity on a screen normal to the beam—this is indeed consistent with the diffraction pattern

Thus

$$\Delta x \, \Delta p_x \sim h \tag{2 7-5}$$

where we have used the de Broglie relation [see Eq (2 2-1)] As can be easily seen, if we make d (and hence Δx) smaller, there will be greater diffraction and hence a greater value of Δp_x It should be noted that Δp_x is almost zero *before* it enters the slit but not *after* it has entered the slit

In order to make an exact statement of the uncertainty principle it is necessary to give precise definitions of Δx and Δp_x These definitions in analogy with the standard deviation in statistics, are

$$\Delta x = \langle (x - \langle x \rangle)^2 \rangle^{1/2} = [\langle x^2 \rangle - \langle x \rangle^2]^{1/2} \tag{2 7-6}$$

and

$$\Delta p_x = \langle (p_x - \langle p_x \rangle)^2 \rangle^{1/2} = [\langle p_x^2 \rangle - \langle p_x \rangle^2]^{1/2} \tag{2 7-7}$$

where

$$\langle x \rangle = \int \Psi^* x \Psi \, d\tau, \text{ etc} \tag{2 7-8}$$

One can then show [see, e g , Ghatak and Lokanathan (1977)]

$$\Delta x \, \Delta p_x \geqslant \tfrac{1}{2}\hbar \tag{2 7-9}$$

the minimum uncertainty ($=\frac{1}{2}\hbar$) occurring for the Gaussian wave packet (see Section 2 2)

For the harmonic oscillator wave functions (see Section 2 8) if we use the various properties of the Hermite–Gauss functions, we get†

$$\langle x \rangle = \int \Psi_n^* x \Psi_n \, dx = 0 \tag{2 7-10}$$

$$\langle x^2 \rangle = \int \Psi_n^* x^2 \Psi_n \, dx = \frac{\hbar}{m\omega}(n + \tfrac{1}{2}) \tag{2 7-11}$$

$$\langle p_x \rangle = \int \Psi_n^* \left(-i\hbar \frac{\partial \Psi_n}{\partial x}\right) dx = 0 \tag{2 7-12}$$

$$\langle p_x^2 \rangle = \int \Psi_n^* \left(-\hbar^2 \frac{\partial^2 \Psi_n}{\partial x^2}\right) dx = m\omega\hbar(n + \tfrac{1}{2}) \tag{2 7-13}$$

Thus

$$\Delta x \, \Delta p_x = (n + \tfrac{1}{2})\hbar \tag{2 7-14}$$

the minimum uncertainty product occurring for the ground state ($n = 0$)

† These relations have been derived in Section 7 6-2 by using the Dirac bra and ket algebra

2.8 The Linear Harmonic Oscillator

In this section we solve the one-dimensional Schrodinger equation [Eq (2 3-23)] for the linear harmonic oscillator problem for which the Hamiltonian is given by

$$H = \frac{p^2}{2m} + \tfrac{1}{2}m\omega^2 x^2 \qquad (2\ 8\text{-}1)$$

where m is the mass of the particle and ω the angular frequency of the classical oscillator The harmonic oscillator problem is of extreme importance in quantum theory as it forms the basis of many topics like the theory of radiation (see Chapter 8), the quantization of lattice vibrations, etc

Using the operator representation of p [see Eq (2 3-8)], the Schrodinger equation (Eq 2 3-11) becomes

$$-\frac{\hbar^2}{2m}\frac{d^2\psi}{dx^2} + \tfrac{1}{2}m\omega^2 x^2\psi = E\psi \qquad (2\ 8\text{-}2)$$

We wish to solve the above equation subject to the boundary condition that ψ must be single valued and finite everywhere†

It is convenient to introduce the dimensionless variables

$$\xi = \left(\frac{m\omega}{\hbar}\right)^{1/2} x, \qquad \varepsilon = \frac{2E}{\hbar\omega} \qquad (2\ 8\text{-}3)$$

Equation (2 8-2) then becomes

$$\frac{d^2\psi}{d\xi^2} + (\varepsilon - \xi^2)\psi(\xi) = 0 \qquad (2\ 8\text{-}4)$$

We next define $u(\xi)$ through the relation

$$\psi(\xi) = \exp(-\tfrac{1}{2}\xi^2)u(\xi) \qquad (2\ 8\text{-}5)$$

Substituting in Eq (2 8-4) we obtain

$$\frac{d^2u}{d\xi^2} - 2\xi\frac{du}{d\xi} + (\varepsilon - 1)u(\xi) = 0 \qquad (2\ 8\text{-}6)$$

† This follows from the fact that since ψ is to be interpreted as the probability amplitude ψ should be single valued at every point in space and $|\psi|\ d\tau$ has to be finite for finite volumes $d\tau$ In some texts the continuity of $\nabla\psi$ is taken as an additional axiom this is not correct In fact the continuity of ψ and $\nabla\psi$ follows from the Schrodinger equation itself as long as V has finite discontinuities indeed if V has infinite discontinuities $\nabla\psi$ is not continuous (see Section 2 11)

We seek a solution of this equation in the form of a power series

$$u(\xi) = \xi^s(a_0 + a_1\xi + a_2\xi^2 + \quad) = \sum_{r=0}^{\infty} a_r \xi^{r+s} \tag{2 8-7}$$

with $a_0 \neq 0$ and $s \geqslant 0$ Differentiating Eq (2 8-7) term by term and substituting in Eq (2 8-6), we obtain

$$\sum_{r=0}^{\infty} (r + s)(r + s - 1)a_r\xi^{r+s-2} - \{2(r + s) - (\varepsilon - 1)\}a_r\xi^{r+s} = 0 \tag{2 8-8}$$

Since the above equation is valid for all values of ξ, we equate the coefficients of each power to zero Thus

$$s(s - 1)a_0 = 0 \tag{2 8-9}$$

$$s(s + 1)a_1 = 0 \tag{2 8-10}$$

and

$$(r + s + 2)(r + s + 1)a_{r+2} = (2r + 2s - \varepsilon + 1)a_r \tag{2 8-11}$$

where

$$r = 0, 1, 2, \tag{2 8-12}$$

Since $a_0 \neq 0$, Eq (2 8-9) (which is known as the indicial equation) gives

$$s = 0 \quad \text{or} \quad s = 1 \tag{2 8-13}$$

Further, Eq (2 8-11) may be rewritten in the form

$$\frac{a_{r+2}}{a_r} = \frac{2r + 2s + 1 - \varepsilon}{(r + s + 2)(r + s + 1)} \rightarrow \frac{2}{r} \quad \text{as} \quad r \rightarrow \infty \tag{2 8-14}$$

The ratio is the same as that of the coefficients of ξ^{r+2} and ξ^r in the expansion† of $\exp(\xi^2)$ This means that if the series [Eq (2 8-7)] is allowed to run to an unlimited number of terms, the wave function [Eq (2 8-5)] would behave as $\exp(\xi^2/2)$ for large ξ and would diverge as $\xi \rightarrow \pm\infty$ This is physically impossible However, this behavior can be averted if beyond a certain coefficient a_r the series is terminated, we can

† In the expansion of $\exp(\xi^2)$ the ratio of the coefficients of ξ^{n+2} and ξ^n is $2/(n + 2)$ n is assumed to be even It should be noted that the infinite series [Eq (2 8-7)] is always convergent It only behaves as $\exp(\xi^2)$ for large values of ξ

do this by setting $a_1 = 0$ (and therefore $a_3 = a_5 = \cdots = 0$)† and by putting

$$2r + 2s + 1 - \varepsilon = 0 \tag{2.8-15}$$

Since $s = 0$ or 1, we have

$$\varepsilon = 2n + 1, \qquad n = 0, 1, 2, \ldots \tag{2.8-16}$$

or

$$E = (n + \tfrac{1}{2})\hbar\omega \tag{2.8-17}$$

Thus the energy eigenvalues are $\frac{1}{2}\hbar\omega$, $\frac{3}{2}\hbar\omega$, $\frac{5}{2}\hbar\omega$, ..., there is only a discrete set—no continuous spectrum exists‡ In order to calculate the corresponding eigenfunction, we illustrate by choosing a specific value of ε say

$$\varepsilon = 5, \qquad \text{i.e.}, n = 2 \tag{2.8-18}$$

We first assume $s = 0$ thus

$$\frac{a_{r+2}}{a_r} = \frac{2r - 4}{(r + 2)(r + 1)} \tag{2.8-19}$$

Obviously $a_2 = -2a_0$ and $a_4 = a_6 = \cdots = 0$ giving

$$u(\xi) = a_0(1 - 2\xi^2) \tag{2.8-20}$$

For $s = 1$

$$\frac{a_{r+2}}{a_r} = \frac{2r - 2}{(r + 3)(r + 1)} \tag{2.8-21}$$

and the even series (see the first footnote on this page) will never terminate, in fact this infinite series will be the other solution of Eq (2.8-6)

† The series given by Eq (2.8-7) can be written in the form

$$u = \xi^s[E + O]$$

where $E = a_0 + a_2\xi^2 + a_4\xi^4 + \cdots$ represents the even power series and $O = a_1\xi + a_3\xi^3 + \cdots$ represents the odd power series Obviously by choosing a particular value of ε [Eq (2.8-16)] both the series cannot be terminated simultaneously [see Eq (2.8-14)], and since we have assumed $a_0 \neq 0$ we must assume $a_1 = 0$ To sum up, for a given odd value of ε one of the two solutions of Eq (2.8-6) is a polynomial and the other is an infinite series which behaves as $\exp(\xi^2)$ for large values of ξ for $\varepsilon \neq 1, 3, 5, \ldots$ both the solutions are infinite series

‡ This is a characteristic of a potential which flares up to ∞ as $x \to \pm\infty$ On the other hand, for a problem like that of the hydrogen atom for which $V(r) = -e^2/(4\pi\varepsilon_0 r)$ (here e is the electronic charge and r is the spherical radial coordinate) an infinite number of discrete states exist for which $E < 0$ and there is a continuous spectrum of eigenvalues in the region $0 < E < \infty$

Using Eq (2 8-16), we can rewrite Eq (2 8-6) as

$$\frac{d^2u}{d\xi^2} - 2\xi\frac{du}{d\xi} + 2nu(\xi) = 0 \qquad (2\ 8\text{-}22)$$

For each integral value of n, this equation has a polynomial solution (as shown above) which is denoted by $H_n(\xi)$ It can be shown that

$$H_n(\xi) = (-1)^n \exp(\xi^2)\frac{d^n}{d\xi^n}\exp(-\xi^2) \qquad (2\ 8\text{-}23)$$

The first few Hermite polynomials are

$$H_0(\xi) = 1 \qquad H_1(\xi) = 2\xi$$
$$H_2(\xi) = 4\xi^2 - 2, \qquad H_3(\xi) = 8\xi^3 - 12\xi, \qquad (2\ 8\text{-}24)$$

Notice that $u(\xi)$ as given by Eq (2 8-20) is a multiple of $H_3(\xi)$ Using Eqs (2 8-5) and (2 8-23), the normalized harmonic oscillator wave functions (i e , $\int_{-\infty}^{+\infty} |\psi_n|^2\, dx = 1$) are

$$\psi_n(x) = N_n \exp\left(-\frac{\alpha^2x^2}{2}\right)H_n(\alpha x) \qquad (2\ 8\text{-}25)$$

$$N_n = \left(\frac{\alpha}{\pi^{1/2}2^n n!}\right)^{1/2} \qquad \xi = \alpha x = \left(\frac{m\omega}{\hbar}\right)^{1/2} x \qquad (2\ 8\text{-}26)$$

2 9 Orthogonality of Wave Functions

We shall first prove that all values of E_n [see Eq (2 3-25)] are real and that if $E_n \neq E_k$ then the corresponding wave functions are necessarily orthogonal i e ,

$$\int \psi_k^* \psi_n \, d\tau = 0 \qquad (2\ 9\text{-}1)$$

We start with the Schrodinger equation for the two states

$$-\frac{\hbar^2}{2m}\nabla^2\psi_n + V(\mathbf{r})\psi_n = E_n\psi_n \qquad (2\ 9\text{-}2)$$

$$-\frac{\hbar^2}{2m}\nabla^2\psi_k + V(\mathbf{r})\psi_k = E_k\psi_k \qquad (2\ 9\text{-}3)$$

We multiply Eq (2 9-2) by ψ_k^* and the complex conjugate of Eq (2 9-3) by ψ_n and subtract

$$-\frac{\hbar^2}{2m}(\psi_k^*\nabla^2\psi_n - \psi_n\nabla^2\psi_k^*) = (E_n - E_k^*)\psi_k^*\psi_n \qquad (2\ 9\text{-}4)$$

or

$$-\frac{\hbar^2}{2m}\int \nabla\cdot(\psi_k^*\nabla\psi_n - \psi_n\nabla\psi_k^*)\,d\tau = (E_n - E_k^*)\int \psi_k^*\psi_n\,d\tau \quad (2.9\text{-}5)$$

Using the divergence theorem†, the integral on the left-hand side can be transformed to a surface integral which would vanish if the volume integral is over the entire space, this is because the wave functions vanish at the surface which is at infinity. Thus

$$(E_n - E_k^*)\int \psi_k^*\psi_n\,d\tau = 0 \quad (2.9\text{-}6)$$

For $n = k$, we must have

$$E_n = E_n^* \quad (2.9\text{-}7)$$

proving that all eigenvalues must be real, and for $E_n \neq E_k$ Eq. (2.9-1) follows.

If $E_n = E_k$ $(n \neq k)$ so that ψ_k and ψ_n are two linearly independent wave functions belonging to the same energy level, then ψ_k and ψ_n are not necessarily orthogonal. An energy level E is said to be degenerate when two or more linearly independent eigenfunctions correspond to it. However, it can easily be shown that any linear combination of the degenerate eigenfunctions (like $C_1\psi_k + C_2\psi_n$) is also a possible eigenfunction belonging to the same eigenvalue:

$$H\psi_k = E_k\psi_k \quad (2.9\text{-}8)$$

$$H\psi_n = E_k\psi_n \quad (2.9\text{-}9)$$

where $H = -(\hbar^2/2m)\nabla^2 + V(\mathbf{r})$ [see Eq. (2.9-2)]. If we multiply Eq. (2.9-8) by C_1 and Eq. (2.9-9) by C_2, where C_1 and C_2 are any complex numbers, and then add we obtain

$$H\phi = E_k\phi \quad (2.9\text{-}10)$$

where $\phi = C_1\psi_k + C_2\psi_n$. Equation (2.9-10) tells us that ϕ is also an eigenfunction belonging to the same eigenvalue. Since C_1 and C_2 are arbitrary, it is always possible to construct linearly independent wave functions (belonging to this level) which are mutually orthogonal. Further,

† According to the divergence theorem

$$\int_V \nabla\cdot\mathbf{F}\,d\tau = \int_S \mathbf{F}\cdot d\mathbf{S}$$

where S is the surface bounding the volume V.

one can always multiply an eigenfunction by a suitable constant such that

$$\int \psi_k^* \psi_n \, d\tau = \delta_{kn} \tag{2 9-11}$$

where δ_{kn} is known as the Kronecker delta function defined through the equation

$$\delta_{kn} = \begin{cases} 0 & \text{if} \quad k \neq n \\ 1 & \text{if} \quad k = n \end{cases} \tag{2 9-12}$$

It may be pointed out that the linear harmonic oscillator states [Eq (2 8-25)] are nondegenerate, however, for the hydrogen atom problem, the state characterized by the quantum number n is n^2-fold degenerate

2 10 The Hydrogenlike Atom Problem

The hydrogenlike atom consists of a nucleus of charge Ze and an electron of charge e Thus the potential energy is given by

$$V(r) = -\frac{Ze^2}{4\pi\epsilon_0 r} \tag{2 10-1}$$

where we have used the MKS system of units and ϵ_0 is the dielectric permittivity of free space The Schrodinger equation (assuming no motion of the nucleus) will be [cf Eq (2 3-22)]

$$\nabla^2\psi + \frac{2\mu}{\hbar^2}\left(E + \frac{Ze^2}{4\pi\epsilon_0 r}\right)\psi = 0 \tag{2 10-2}$$

where μ is the mass of the electron† One can solve the above equation by using the method of separation of variables‡ and then each component equation by using a method similar to that used in the harmonic oscillator problem The mathematics is quite involved [see, e g , Ghatak and Lokanathan (1977)], the results are given by Eqs (2 10-3) to (2 10-22)

† In the rigorous theory if we do take into account the motion of the nucleus then μ would represent the reduced mass i e

$$\mu = \frac{m_e m_N}{m_e + m_N}$$

where m_e and m_N represent the masses of the electron and nucleus respectively Since $m_N \gg m_e$ $\mu \simeq m_e$

‡ The spherical polar coordinates have to be used for which

$$\nabla^2\psi = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial\psi}{\partial r}\right) + \frac{1}{r^2 \sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial\psi}{\partial\theta}\right) + \frac{1}{r^2\sin^2\theta}\frac{\partial^2\psi}{\partial\varphi^2}$$

First, the energy eigenvalues are given by

$$E_n = -\frac{\mu Z^2}{2\hbar^2 n^2}\left(\frac{e^2}{4\pi\epsilon_0}\right)^2 \simeq -\frac{Z^2}{n^2}\,13\,6\,\text{eV} \tag{2 10-3}$$

where $n = 1, 2\ 3$ represents the total quantum number The corresponding eigenfunctions are given by

$$\psi_{nlm}(r, \theta, \varphi) = R_{nl}(r)Y_{lm}(\theta, \varphi) \tag{2 10-4}$$

where $R_{nl}(r)$ represent the radial part of the wave function and $Y_{lm}(\theta, \varphi)$ represent the angular part of the wave function and are known as the spherical harmonics For a given value of n

$$l = 0, 1, 2, \quad , (n-1) \tag{2 10-5}$$

and for each value of l

$$m = -l, -l+1, \quad , l-1, l \tag{2 10-6}$$

The first few eigenfunctions are given below†

$$R_{10} = 2\left(\frac{Z}{a_0}\right)^{3/2} e^{-\xi} \tag{2 10-7}$$

$$R_{20} = \frac{1}{2^{1/2}}\left(\frac{Z}{a_0}\right)^{3/2} e^{-\xi/2}\left(1 - \frac{\xi}{2}\right) \tag{2 10-8}$$

$$R_{21} = \frac{1}{2(6^{1/2})}\left(\frac{Z}{a_0}\right)^{3/2} e^{-\xi/2}\xi \tag{2 10-9}$$

$$R_{30} = \frac{2}{3(3^{1/2})}\left(\frac{Z}{a_0}\right)^{3/2} e^{-\xi/3}\left(1 - \frac{2\xi}{3} + \frac{2\xi^2}{27}\right) \tag{2 10-10}$$

$$R_{31} = \frac{8}{27(6^{1/2})}\left(\frac{Z}{a_0}\right)^{3/2} e^{-\xi/3}\left(\xi - \frac{\xi^2}{6}\right) \tag{2 10-11}$$

$$R_{32} = \frac{4}{81(30^{1/2})}\left(\frac{Z}{a_0}\right)^{3/2} \xi^2 e^{-\xi/3} \tag{2 10-12}$$

etc , in the above equations $\xi = Zr/a_0$ and

$$a_0 = \frac{\hbar^2}{\mu}\left(\frac{4\pi\epsilon_0}{e^2}\right) \simeq \tfrac{1}{2}\,\text{Å} = 0\,5 \times 10^{-10}\,\text{m} \tag{2 10-13}$$

† Unfortunately different authors use different signs etc in the definition of the wave function We follow Condon and Shortley (1935)

is known as the Bohr radius Further,

$$Y_{0\,0} = (4\pi)^{-1/2} \tag{2 10-14}$$

$$Y_{1\,1} = -\left(\frac{3}{8\pi}\right)^{1/2} \sin\theta e^{i\varphi} \tag{2 10-15}$$

$$Y_{1\,0} = \left(\frac{3}{4\pi}\right)^{1/2} \cos\theta \tag{2 10-16}$$

$$Y_{1\,-1} = \left(\frac{3}{8\pi}\right)^{1/2} \sin\theta e^{-i\varphi} \tag{2 10-17}$$

$$Y_{2\,2} = \left(\frac{15}{32\pi}\right)^{1/2} \sin^2\theta e^{2i\varphi} \tag{2 10-18}$$

$$Y_{2\,1} = -\left(\frac{15}{8\pi}\right)^{1/2} \sin\theta\cos\theta e^{i\varphi} \tag{2 10-19}$$

$$Y_{2\,0} = \left(\frac{5}{16\pi}\right)^{1/2} (3\cos^2\theta - 1) \tag{2 10-20}$$

$$Y_{2\,-1} = \left(\frac{15}{8\pi}\right)^{1/2} \sin\theta\cos\theta e^{-i\varphi} \tag{2 10-21}$$

$$Y_{2\,-2} = \left(\frac{15}{32\pi}\right)^{1/2} \sin^2\theta e^{-2i\varphi} \tag{2 10-22}$$

etc The ground state eigenfunction is $\psi_{1\,0\,0}$ ($n = 1$ $l = 0$, $m = 0$), the first excited state ($n = 2$) is fourfold degenerate $\psi_{2\,0\,0}$, $\psi_{2\,1\,-1}$, $\psi_{2\,1\,0}$, and $\psi_{2\,1\,1}$ Similarly the $n = 3$ state is ninefold degenerate In general, the states characterized by the quantum number n are n^2-fold degenerate The wave functions are orthonormal, i e

$$\iiint \psi^*_{nlm}\psi_{n'l'm'}\, r^2\, dr \sin\theta\, d\theta\, d\varphi = \delta_{nn'}\,\delta_{ll'}\,\delta_{mm'} \tag{2 10-23}$$

It may be a worthwhile exercise for the reader to see that the above wave functions indeed satisfy Eq (2 10-2) with E given by Eq (2 10-3)

2 11 *Some Simple Solutions of the Schrodinger Equation*

We will briefly discuss below some potential energy distributions for which it is possible to obtain exact solutions of the Schrodinger equation†

† The detailed derivations of the solutions are quite straightforward and are left as an exercise to the reader

Most of these distributions are of importance in solid state physics, atomic and molecular physics, etc The first two cases involve solution of the one-dimensional Schrodinger equation

(i) For

$$V(x) = \begin{cases} 0, & 0 < x < L \\ \infty, & x < 0 \text{ and } x > L \end{cases} \tag{2 11-1}$$

$$\psi_n(x) = \left(\frac{2}{L}\right)^{1/2} \sin\left(\frac{n\pi}{L}x\right), \qquad 0 < x < L \text{ and } 0 \text{ everywhere else} \tag{2 11-2}$$

$$E_n = \frac{\pi^2\hbar^2}{2\mu L^2}n^2, \qquad n = 1, 2, 3, \tag{2 11-3}$$

Since there is an infinite barrier in potential energy, $\psi(x) = 0$ at $x = 0$ and at $x = L$ These are the boundary conditions which lead to the above solutions Notice only discrete states exist and $d\psi/dx$ is *not* continuous at $x = 0$ and $x = L$

(ii) For

$$V(x) = \begin{cases} -V_0, & -a < x < a \\ 0, & |x| > a \end{cases} \tag{2 11-4}$$

assuming continuity of ψ and $d\psi/dx$ everywhere, one obtains the following transcendental equations which determine the discrete energy levels $(E < 0)$

$$\left(\frac{-E}{V_0 + E}\right)^{1/2} = \tan\left\{\frac{a}{\hbar}[2\mu(V_0 + E)]^{1/2}\right\} \tag{2 11-5}$$

for the symmetric wave functions $[\psi(x) = \psi(-x)]$ and

$$\left(\frac{-E}{V_0 + E}\right)^{1/2} = -\cot\left\{\frac{a}{\hbar}[2\mu(V_0 + E)]^{1/2}\right\} \tag{2 11-6}$$

for the antisymmetric wave functions $[\psi(x) = -\psi(-x)]$ The corresponding wave functions are in terms of sine and cosine functions in the region $|x| < a$ and in terms of exponentials in the region $|x| > a$ It may be noted that for discrete states $E < 0$, for $E > 0$, the energy states form a continuum

(iii) For a particle confined in a cube we may write

$$V(x, y, z) = \begin{cases} 0 & \text{for } 0 < x < L, \quad 0 < y < L, \quad 0 < z < L \\ \infty & \text{everywhere else} \end{cases}$$

Using the method of separation of variables and assuming ψ to vanish on the surface of the cube, one gets

$$\psi = \left(\frac{8}{L^3}\right)^{1/2} \sin\left(\frac{n_x\pi}{L}x\right)\sin\left(\frac{n_y\pi}{L}y\right)\sin\left(\frac{n_z\pi}{L}z\right) \qquad (2\ 11\text{-}7)$$

$$E_{n_x n_y n_z} = \frac{\pi^2\hbar^2}{2\mu L^2}(n_x^2 + n_y^2 + n_z^2) \qquad (2\ 11\text{-}8)$$

$$n_x, n_y, n_z = 1, 2, 3, \qquad (2\ 11\text{-}9)$$

(iv) For the three-dimensional harmonic oscillator

$$V = \tfrac{1}{2}\mu(\omega_1^2 x^2 + \omega_2^2 y^2 + \omega_3^2 z^2) \qquad (2\ 11\text{-}10)$$

one can solve the Schrodinger equation by using the method of separation of variables The energy levels are

$$E = (n_1 + \tfrac{1}{2})\hbar\omega_1 + (n_2 + \tfrac{1}{2})\hbar\omega_2 + (n_3 + \tfrac{1}{2})\hbar\omega_3 \qquad (2\ 11\text{-}11)$$

$n_1, n_2, n_3 = 0, 1, 2,$ The corresponding wave functions are products of the Hermite–Gauss functions

13

The Einstein Coefficients and Light Amplification

3.1 Introduction

As mentioned in Chapter 1, Einstein in 1917 predicted that one may have two kinds of emissions, one he called spontaneous emission and the other stimulated emission, which is caused by the presence of light radiation of the proper frequency. In Section 3.2 we define the Einstein coefficients governing these processes and give the original argument of Einstein which led to a relation between these coefficients. We also show how a light beam gets amplified in the presence of population inversion.

In Section 3.3, we give the quantum theory for the calculation of the transition rates and also obtain explicit expressions for the Einstein coefficients. In Section 3.4 we obtain more exact solutions of the Schrodinger equation for a two-state system. In Section 3.5, we discuss the origin of the broadening of spectral lines.

3.2 The Einstein Coefficients

Let N_1 and N_2 represent the number of atoms per unit volume in levels 1 and 2, respectively, the levels correspond to energies E_1 and E_2 (see Fig. 3.1). An atom in the lower energy level can absorb radiation and get excited to the level E_2. This excitation process can occur only in the presence of radiation. Such a process is called absorption. The rate of absorption would depend on the density of radiation at the particular frequency separating the two levels. Thus, if

$$\omega = \frac{E_2 - E_1}{\hbar} \qquad (3.2\text{-}1)$$

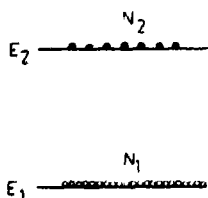



Fig 3 1 N_1 and N_2 represent the number of atoms in levels 1 and 2 with energies E_1 and E_2 respectively • Excited atoms ∘ atoms in the ground state

then the absorption process depends on the energy density of radiation at the frequency ω, this energy density is denoted by $u(\omega)$ and is defined such that $u(\omega)\, d\omega$ represents the radiation energy per unit volume within the frequency interval ω and $\omega + d\omega$ The rate of absorption would be proportional to N_1 and also to $u(\omega)$ Thus, the number of absorptions per unit time per unit volume can be written as

$$N_1 B_{12} u(\omega) \tag{3 2-2}$$

where B_{12} is the coefficient of proportionality and is a characteristic of the energy levels

Let us now consider the reverse process, namely, the emission of radiation at a frequency ω when the atom deexcites from the level E_2 to E_1 As mentioned in Chapter 1, Einstein postulated that an atom in an excited level can make a radiative transition to a lower energy level either through spontaneous emission or through stimulated emission In spontaneous emission, the probability per unit time of the atom making a downward transition is independent of the energy density of the radiation field and depends only on the levels involved in the transition If we represent the coefficient of proportionality by A_{21} then

$$N_2 A_{21} \tag{3 2-3}$$

would represent the rate of spontaneous emissions to the lower energy level

In the case of stimulated emission the rate of transition to the lower energy level is directly proportional to the energy density of the radiation at the frequency ω Thus the rate of stimulated emissions would be given by

$$N_2 B_{21} u(\omega) \tag{3 2-4}$$

The quantities A_{21}, B_{12}, and B_{21} are known as Einstein coefficients and are determined by the atomic system

At thermal equilibrium, the number of upward transitions must be equal to the number of downward transitions Thus, we may write (at

thermal equilibrium)

$$N_1 B_{12} u(\omega) = N_2 A_{21} + N_2 B_{21} u(\omega)$$

or

$$u(\omega) = \frac{A_{21}}{(N_1/N_2)B_{12} - B_{21}} \tag{3.2-5}$$

From Boltzmann's law, we have the following expressions for the ratio of the populations of two levels at temperature T

$$\frac{N_1}{N_2} = \exp\left(\frac{E_2 - E_1}{k_B T}\right) = \exp\left(\frac{\hbar\omega}{k_B T}\right) \tag{3.2-6}$$

where k_B represents the Boltzmann constant. Thus, we may write

$$u(\omega) = \frac{A_{21}}{B_{12}\exp(\hbar\omega/k_B T) - B_{21}} \tag{3.2-7}$$

Now, according to Planck's law† the energy density of radiation is given by

$$u(\omega) = \frac{\hbar\omega^3}{\pi^2 c^3}\frac{1}{\exp(\hbar\omega/k_B T) - 1} \tag{3.2-8}$$

Comparing Eqs. (3.2-7) and (3.2-8), we obtain‡

$$B_{12} = B_{21} = B \tag{3.2-9}$$

and

$$\frac{A_{21}}{B_{21}} = \frac{\hbar\omega^3}{\pi^2 c^3} \tag{3.2-10}$$

Thus, the probabilities of stimulated absorption and stimulated emission are the same and the ratio of the A and B coefficients is given by Eq. (3.2-10). At thermal equilibrium, the ratio of the number of spontaneous to stimulated emissions is given by

$$\frac{A}{Bu(\omega)} = \exp\left(\frac{\hbar\omega}{k_B T}\right) - 1 \tag{3.2-11}$$

where we have used Eqs. (3.2-8) and (3.2-10). Hence, at thermal equilibrium (at temperature T), for $\omega \ll k_B T/\hbar$, the number of stimulated emissions far exceeds the number of spontaneous emissions while for

† See Appendix C.

‡ If the levels 1 and 2 are g_1- and g_2-fold degenerate, then $N_1/N_2 = g_1/g_2 \exp(\hbar\omega/k_B T)$, $B_{12} = B_{21}(g_2/g_1)$ and $A_{21}/B_{21} = \hbar\omega^3/\pi^2 c^3$.

$\omega \gg k_B T/\hbar$, the number of spontaneous emissions far exceeds the number of stimulated emissions For normal optical sources, $T \sim 10^3$ °K and

$$\frac{k_B T}{\hbar} \simeq \frac{1.38 \times 10^{-23}\,(\mathrm{J/°K}) \times 10^3\,(\mathrm{°K})}{1.054 \times 10^{-34}\,(\mathrm{J\,sec})} \simeq 1.3 \times 10^{14}\,\mathrm{sec}^{-1}$$

Since for the optical region $\omega \sim 4 \times 10^{15}\,\mathrm{sec}^{-1}$, we find that at optical frequencies the emission is predominantly due to spontaneous transitions and hence the emission from usual light sources is incoherent

In the above discussion we have assumed that the atom is capable of interacting with radiation of a particular frequency ω, however, in general, the atom can interact with radiation over a range of frequencies and the strength of the interaction is a function of the frequency This function is known as the line-shape function Let $g(\omega)$ represent the normalized line-shape function† corresponding to the transition between levels 1 and 2, then $n_{1\omega}\,d\omega$ $[= N_1 g(\omega)\,d\omega]$ and $n_{2\omega}\,d\omega$ $[= N_2 g(\omega)\,d\omega]$ would represent the number of atoms per unit volume in levels 1 and 2, respectively, which are capable of interacting with radiation of frequency between ω and $\omega + d\omega$ Now,

$$B_{21} = \frac{\pi^2 c^3}{\hbar\omega^3} A_{21} = \frac{\pi^2 c^3}{\hbar\omega^3 t_{sp}} \qquad (3.2\text{-}12)$$

where t_{sp} $(\equiv 1/A_{21})$ represents the spontaneous lifetime of the upper level Thus if we take into account the line-shape function, the total number of stimulated emissions per unit time per unit volume would be given by

$$\begin{aligned} W_{21} &= N_2 \int B_{21} u(\omega) g(\omega)\,d\omega \\ &= N_2 \frac{\pi^2 c^3}{\hbar t_{sp}} \int \frac{u(\omega)}{\omega^3} g(\omega)\,d\omega \end{aligned} \qquad (3.2\text{-}13)$$

If $u(\omega)/\omega^3$ can be assumed to be essentially constant over the region where $g(\omega)$ is appreciable, one immediately obtains

$$W_{21} \simeq N_2 B_{21} u(\omega) = \frac{N_2 \pi^2 c^3}{\hbar\omega^3 t_{sp}} u(\omega) \qquad (3.2\text{-}14)$$

† The line-shape function will be discussed in detail in Section 3.5 the normalization condition is

$$\int g(\omega)\,d\omega = 1$$

consistent with Eq (3 2-4) However for a near-monochromatic radiation field† at frequency ω' we have

$$u(\omega) = u_\omega \delta(\omega - \omega') \tag{3 2-15}$$

where u_ω represents the energy density associated with the monochromatic field‡ For such a case, Eq (3 2-13) gives

$$W_{21} = N_2 \frac{\pi^2 c^3}{\hbar t_{sp}} \frac{u_\omega}{\omega'^3} g(\omega') \tag{3 2-16}$$

where $g(\omega')$ represents the value of the line-shape function evaluated at the radiation frequency ω' Similarly, the number of stimulated absorptions per unit time per unit volume would be given by

$$W_{12} = N_1 \frac{\pi^2 c^3}{\hbar t_{sp}} \frac{u_\omega}{\omega'^3} g(\omega') \tag{3 2-17}$$

We next consider a collection of atoms and let a near monochromatic beam of frequency ω be propagating through it along the z direction In order to obtain an expression for the rate of change of the intensity of the beam as it propagates we consider two planes of area S perpendicular to the z direction at z and $z + dz$ (see Fig 3 2) The volume of the medium between planes P_1 and P_2 is $S\,dz$ and hence the number of stimulated absorptions per unit time would be $W_{12}S\,dz$ Since each photon has an

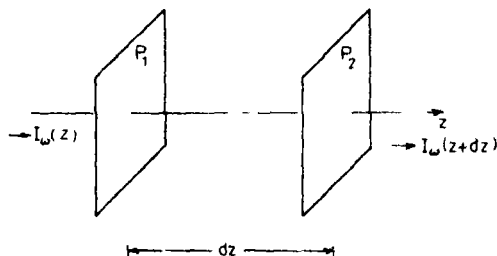



Fig 3 2 Radiation propagating along the z direction through a collection of atoms

energy $\hbar\omega$, the energy absorbed per unit time in the volume element $S\,dz$ would be

$$W_{12}\hbar\omega S\,dz$$

Similarly the corresponding energy gain (because of stimulated emissions)

† By near-monochromatic radiation field we imply that the width of the function $u(\omega)$ is extremely small in comparison to the width of the line-shape function

‡ It may be noted that the units of $u(\omega)$ and u_ω are different $u(\omega)$ has units of energy per unit volume per unit frequency interval while u_ω is simply energy per unit volume

L—4

would be

$$W_{21}\hbar\omega S\,dz$$

where we have neglected the radiation arising out of spontaneous emissions, because such radiations propagate in random directions and are, in general, lost from the beam Thus, the net amount of energy absorbed per unit time in the volume element $S\,dz$ and in the frequency interval ω and $\omega + d\omega$ would be

$$(W_{12} - W_{21})\hbar\omega S\,dz$$

If $I_\omega(z)$ represents the intensity of the beam in the plane P_1, then the total energy entering the volume element $S\,dz$ per unit time will be

$$I_\omega(z)S$$

Similarly, if $I_\omega(z + dz)$ represents the intensity in the plane P_2 then the total energy leaving the volume element per unit time would be

$$I_\omega(z + dz)S = I_\omega(z)S + \frac{\partial I_\omega}{\partial z}\,dz\,S \qquad (3\,2\text{-}18)$$

Hence the net amount of energy leaving the volume element per unit time would be

$$\frac{\partial I_\omega}{\partial z}\,dz\,S$$

This must be equal to the negative of the net energy absorbed by the medium between z and $z + dz$ Thus,

$$\frac{\partial I_\omega}{\partial z}\,S\,dz = -(W_{12} - W_{21})\hbar\omega S\,dz$$

$$= -\frac{\pi^2 c^3}{\hbar t_{sp}\omega^3}\,g(\omega)u_\omega \hbar\omega S\,dz\,(N_1 - N_2)$$

or

$$\frac{\partial I_\omega}{\partial z} = -\frac{\pi^2 c^3}{\omega^2 t_{sp}}\,g(\omega)u_\omega(N_1 - N_2) \qquad (3\,2\text{-}19)$$

Now the energy density u_ω and the intensity I_ω are related through the following equation†

$$I_\omega = v u_\omega = \frac{c}{n_0}\,u_\omega \qquad (3\,2\text{-}20)$$

† This is analogous to the equation $\mathbf{J} = \rho\mathbf{v}$ where ρ represents the number of particles per unit volume (all propagating with velocity $\mathbf{v}$) and $\mathbf{J}$ represents the number of particles crossing a unit area (perpendicular) to the direction of propagation per unit time This can be easily seen from the fact that the number of particles crossing a unit area per unit time would be the particles contained in a cylinder of length v units having unit area of cross section

where v ($= c/n_0$) represents the velocity of the radiation field in the medium, n_0 being its refractive index Thus Eq (3 2-19) reduces to

$$\frac{1}{I_\omega}\frac{dI_\omega}{dz} = -\alpha_\omega \tag{3 2-21}$$

where†

$$\alpha_\omega = \frac{\pi^2 c^2 n_0}{\omega^2 t_{sp}}(N_1 - N_2)g(\omega) \tag{3 2-22}$$

Figure 3 3 gives a typical plot of α_ω with ω Now, Eq (3 2-21) can be integrated to give‡

$$I_\omega(z) = I_\omega(0)\exp(-\alpha_\omega z) \tag{3 2-23}$$

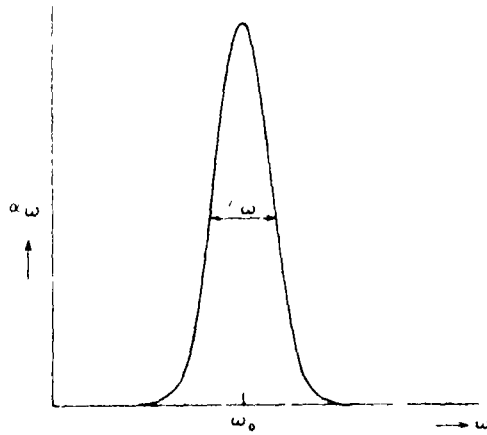



Fig 3 3 A typical plot of α_ω vs ω

Thus, if α_ω is positive, i e , if $N_1 > N_2$, then the intensity of the beam decreases exponentially, with the intensity decreasing to $1/e$ of its value at $z = 0$ in a distance $1/\alpha_\omega$ Hence, at thermal equilibrium, since the number of atoms in the lower level is greater than that in the upper level, the energy of the beam decreases exponentially as it propagates through the medium

† In the presence of degeneracies, the factor $(N_1 - N_2)$ in Eq (3 2-22) is replaced by $(N_1 g_2/g_1 - N_2)$ where the levels 1 and 2 are g_1- and g_2-fold degenerate

‡ In obtaining Eq (3 2-23) from Eq (3 2-21), it has been assumed that $(N_1 - N_2)$ (and hence α_ω) is independent of I_ω Such an approximation is valid only for small values of I_ω When I_ω becomes very large i e for intense light beams, saturation of the levels sets in and then the attenuation is linear rather than exponential This is also discussed in detail in Sections 4 2 and 4 3

On the other hand if there are more atoms in the excited level than in the lower level (i.e. there is a population inversion) then $\alpha_r < 0$ and there would be an exponential increase in intensity of the beam this is known as light amplification

In the actual laser system, the active medium (which is capable of amplification) is placed between a pair of mirrors forming what is known as a resonator (see Chapter 6) In order that oscillations be sustained in the cavity it is essential that the net losses suffered by the beam be compensated by the gain of the medium At threshold and under steady state operation the two are exactly compensated In order to obtain the threshold condition let d be the length of the active medium let R_1 and R_2 represent the energy reflectivities of the mirrors at the two ends of the laser resonator and let I represent the intensity of the beam at one of the mirrors In traveling a length d through the active medium the beam gets amplified by a factor† $e^{-\alpha_r d}$ and gets diminished by a factor $e^{-\alpha_l d}$ due to losses caused by absorption scattering etc in the laser medium On suffering a reflection at the mirror with reflectivity R_1 the intensity reduces to $IR_1 \exp(-\alpha_r d - \alpha_l d)$ A further traversal through a length d and a reflection at the mirror of reflectivity R_2 results in the beam having an intensity $IR_1R_2 \exp[-2d(\alpha_r + \alpha_l)]$ Thus for the laser oscillation to begin

$$R_1R_2 \exp[-2d(\alpha_r + \alpha_l)] \geqslant 1 \qquad (3.2\text{-}24)$$

The equality sign would correspond to the threshold value for oscillation When the laser is oscillating in steady state the total round trip gain must again be equal to unity and the population difference will again be given by Eq (3.2-24) with the equality sign Equation (3.2-24) can be rearranged to give

$$e^{-2\alpha_r d} \geqslant \frac{1}{R_1R_2} e^{2\alpha_l d} \qquad (3.2\text{-}25)$$

The right-hand side of Eq (3.2-25) depends only on the passive cavity parameters In fact it is related to the quality factor Q of the passive resonator From Eq (6.3-12) we obtain‡

$$2\alpha_l d - \ln R_1R_2 = \frac{4\pi\nu n_0 d}{cQ}$$

or

$$\frac{e^{2\alpha_l d}}{R_1R_2} = \exp\left(\frac{4\pi\nu n_0 d}{cQ}\right) \qquad (3.2\text{-}26)$$

† In an amplifying medium $\alpha_r < 0$ and $e^{-\alpha_r d} > 1$

‡ In Section 6.3 we have discussed in greater detail the quality factor associated with a resonator mode

where ν represents the oscillation frequency at the center of the resonator mode Thus from Eqs (3 2-25) and (3 2-26) we obtain

$$-\alpha_\omega \geqslant \frac{2\pi\nu n_0}{cQ} \qquad (3\ 2\text{-}27)$$

Using the value of α_ω given by Eq (3 2-22) we get

$$N_2 - N_1 \geqslant \frac{8\pi\nu^3}{c^3} \frac{1}{Q} \frac{t_{sp}}{g(\omega)} \qquad (3\ 2\text{-}28)$$

The quality factor Q is related to the passive cavity lifetime through Eq (6 3-4) Hence we can also write

$$N_2 - N_1 \geqslant \frac{4\nu^2}{c^3} \frac{1}{g(\omega)} \frac{t_{sp}}{t_c} \qquad (3\ 2\text{-}29)$$

The above equation gives the threshold population inversion required for the oscillation of the laser Clearly, the minimum threshold value would correspond to the center of the line where $g(\omega)$ is a maximum

As the laser medium is pumped harder and harder, the population inversion between the two levels goes on increasing The mode that lies nearest to the resonance frequency of the atomic system reaches threshold first and begins to oscillate As the pumping is still further increased, the nearby modes may also reach threshold and start oscillating

According to Eq (3 2-29), in order to have a low threshold value of population inversion, the following conditions must hold

(i) The value of t_c should be large, i e , the losses in the cavity must be small

(ii) The value of $g(\omega)$ at the center of the line is $2/(\pi\,\Delta\omega)$ for a Lorentzian line and $2(\pi \ln 2)^{1/2}/(\pi\,\Delta\omega)$ for a Gaussian line (see Section 3 5) Thus smaller values of $\Delta\omega$ (the width of the line) lead to smaller values of threshold population inversion

(iii) Smaller values of t_{sp} (i e , strongly allowed transitions) also lead to smaller values of threshold inversion It must be noted here that for shorter relaxation rates, larger pumping power is required to maintain a given amount of population inversion In general, population inversion is more easily obtained on transitions which have longer relaxation times

(iv) The value of $g(\omega)$ at the center of the line is inversely proportional to $\Delta\omega$, which, for example, in the case of Doppler broadening is proportional to ω (see Section 3 5) Thus, the threshold population inversion increases approximately in proportion to ω^3 (apart from the frequency dependence of the other terms) Hence it is much easier to

obtain laser action at the infrared wavelengths as compared to the ultraviolet region.

In order to get an idea of the magnitude of population inversion required for oscillation, we consider a ruby laser.† Let us consider the laser to be oscillating at the frequency corresponding to the peak of the emission line. We assume a concentration of 0.05% of Cr^{3+} ions in the crystal; this corresponds to a population of $N = 1.6 \times 10^{19}$ Cr^{3+} ions/cm³. For the case of ruby, the line is homogeneously broadened‡ and the value of $g(\omega)$ at the peak of the line is $2/(\pi\,\Delta\omega)$; hence the threshold population inversion density would be

$$(N_2 - N_1)_t = \frac{4\pi^2}{\lambda^3}\frac{\Delta\nu}{\nu}\frac{t_{sp}}{t_c} \qquad (3.2\text{-}30)$$

where t_{sp} is the spontaneous relaxation time of the upper laser level and t_c is the cavity lifetime. [For a Gaussian line shape the threshold inversion would be given by Eq. (3.2-30) with $\Delta\nu$ replaced by $\Delta\nu/(\pi \ln 2)^{1/2}$.] For ruby laser transition, one has

$$\begin{aligned} \nu &\simeq 4.3 \times 10^{14}\,\text{sec}^{-1} \qquad \Delta\nu \simeq 1.5 \times 10^{11}\,\text{sec}^{-1} \\ t_{sp} &\simeq 3 \times 10^{-3}\,\text{sec} \\ \lambda &= \frac{0.6943}{1.76} \times 10^{-4} \simeq 4 \times 10^{-5}\,\text{cm} \end{aligned} \qquad (3.2\text{-}31)$$

where we have used the fact that the refractive index of ruby is 1.76.

In order to calculate t_c we note that if d is the length of the optical cavity, n_0 is the refractive index of the medium filling the cavity, and x is the fractional loss per round trip, then the cavity lifetime t_c is given by [see Eq. (6.3-15)]

$$t_c = \frac{2 n_0 d}{c \ln[1/(1 - x)]} \qquad (3.2\text{-}32)$$

Thus if we assume a cavity length of 5 cm and a loss per round trip of 10%, then

$$t_c \simeq 6 \times 10^{-9}\,\text{sec} \qquad (3.2\text{-}33)$$

Substituting all these values in Eq. (3.2-30) we get for the threshold population inversion density

$$(N_2 - N_1)_t \simeq 1.1 \times 10^{16}\,Cr^{3+}\,\text{ions/cm}^3 \qquad (3.2\text{-}34)$$

† The ruby laser is discussed in greater detail in Chapter 9. It is an example of a three level laser system and was the first operating laser. The active medium consists of Cr^{3+} ions which are doped in the host crystal Al_2O_3.

‡ Homogeneous broadening is discussed in Section 3.5.

Since the total density of Cr^{3+} ions in ruby is about 10^{19}, the fractional excess population required is very small

3.3 Quantum Theory for the Evaluation of the Transition Rates and Einstein Coefficients

For the calculation of the transition rates we consider the atom to be in the presence of an oscillating electric field

$$\mathscr{E}(t) = \hat{\mathbf{e}}E_0 \cos \omega t \tag{3 3-1}$$

which is switched on at $t = 0$, $\hat{\mathbf{e}}$ represents the unit vector along the direction of the electric field The frequency ω is assumed to be very close to the resonant frequency $[(E_2 - E_1)/\hbar]$ corresponding to the transition from the state 1 to 2 (see Fig 3 1) We will show that the presence of the higher excited states can be neglected because the corresponding transition frequencies are far away from ω In the presence of the electric field, the time-dependent Schrodinger equation becomes

$$i\hbar \frac{\partial \Psi}{\partial t} = (H_0 + H')\Psi \tag{3 3-2}$$

where

$$H' = -e\mathscr{E} \cdot \mathbf{r} = -eE_0(\hat{\mathbf{e}} \cdot \mathbf{r}) \cos \omega t \tag{3 3-3}$$

represents the interaction energy of the electron with the electric field and H_0 (which is independent of time) represents the Hamiltonian of the atom, e (<0) represents the charge of the electron † Since H_0 is independent of time, the solution of the Schrodinger equation

$$i\hbar \frac{\partial \Psi}{\partial t} = H_0 \Psi \tag{3 3-4}$$

is of the form

$$\Psi = \psi_n(\mathbf{r}) e^{-iE_n t/\hbar} \tag{3 3-5}$$

where $\psi_n(\mathbf{r})$ and E_n are the eigenfunctions and eigenvalues of H_0

$$H_0 \psi_n(\mathbf{r}) = E_n \psi_n(\mathbf{r}) \tag{3 3-6}$$

† We are considering here a single-electron atom with $\mathbf{r}$ representing the position vector of the electron with respect to the nucleus Thus the electric dipole moment of the atom is given by $\mathbf{p} = e\mathbf{r}$ because the direction of the dipole moment is from the negative charge to the positive charge The interaction energy of a dipole placed in an electric field $\mathscr{E}$ is $-\mathbf{p} \cdot \mathscr{E}$, which leads to Eq (3 3-3)

The functions $\psi_n(\mathbf{r})$ are known as the atomic wave functions and satisfy the orthonormality condition

$$\int \psi_n^*(\mathbf{r})\psi_m(\mathbf{r})\,d\tau = \delta_{nm} = \begin{cases} 0 & \text{if } n \neq m \\ 1 & \text{if } n = m \end{cases} \tag{3.3-7}$$

The solution of Eq. (3.3-2) can always be written as a linear combination of the atomic wave functions

$$\Psi(\mathbf{r}, t) = \sum_n C_n(t)\psi_n(\mathbf{r})e^{-i\omega_n t} \tag{3.3-8}$$

where

$$\omega_n = E_n/\hbar \tag{3.3-9}$$

Substituting from Eq. (3.3-8) in Eq. (3.3-2), we obtain

$$i\hbar \sum_n \left(\frac{dC_n}{dt} - i\omega_n C_n\right)e^{-i\omega_n t}\psi_n(\mathbf{r}) = \sum_n E_n C_n(t)\psi_n(\mathbf{r})e^{-i\omega_n t}$$

$$- eE_0(\hat{\mathbf{e}}\cdot\mathbf{r})\sum_n C_n(t)\psi_n(\mathbf{r})e^{-i\omega_n t}\cos\omega t \tag{3.3-10}$$

where we have used Eq. (3.3-6). It is immediately seen that the second term on the left-hand side exactly cancels with the first term on the right-hand side. If we multiply by ψ_m^* and integrate we would get

$$i\hbar\frac{dC_m}{dt} = \tfrac{1}{2}E_0 \sum_n \mathscr{D}_{mn}C_n(t)(e^{i(\omega_{mn}+\omega)t} + e^{i(\omega_{mn}-\omega)t}) \tag{3.3-11}$$

where use has been made of the orthogonality relation [Eq. (3.3-7)] and

$$\omega_{mn} \equiv \omega_m - \omega_n = \frac{E_m - E_n}{\hbar} \tag{3.3-12}$$

$$\mathscr{D}_{mn} = \hat{\mathbf{e}}\cdot\mathbf{P}_{mn} \tag{3.3-13a}$$

$$\mathbf{P}_{mn} = -e\int \psi_m^*(\mathbf{r})\mathbf{r}\psi_n(\mathbf{r})\,d\tau = |e|\int \psi_m^*(\mathbf{r})\mathbf{r}\psi_n(\mathbf{r})\,d\tau \tag{3.3-13b}$$

We wish to solve Eq. (3.3-11) subject to the boundary condition

$$\begin{aligned} C_k(t = 0) &= 1 \\ C_n(t = 0) &= 0 \qquad \text{for } n \neq k \end{aligned} \tag{3.3-14}$$

i.e., at $t = 0$ the atom is assumed to be in the state characterized by the wave function ψ_k. Equation (3.3-11) represents an infinite set of coupled equations and as a first approximation, one may replace $C_n(t)$ by $C_n(0)$

on the right-hand side of Eq (3 3-11) Thus,

$$i\hbar \frac{dC_m}{dt} \simeq \tfrac{1}{2}E_0\mathcal{D}_{mk}\left(e^{i(\omega_{mk}+\omega)t} + e^{i(\omega_{mk}-\omega)t}\right) \qquad (3\ 3\text{-}15)$$

Integrating, one obtains

$$C_m(t) - C_m(0) \simeq -\frac{E_0}{2\hbar}\mathcal{D}_{mk}\left[\frac{e^{i(\omega_{mk}+\omega)t} - 1}{(\omega_{mk} + \omega)} + \frac{e^{i(\omega_{mk}-\omega)t} - 1}{(\omega_{mk} - \omega)}\right]$$

or, for $m \neq k$,

$$C_m(t) = -i\mathcal{D}_{mk}\frac{E_0}{\hbar}\left[e^{i(\omega_{mk}+\omega)t/2}\frac{\sin(\omega_{mk} + \omega)t/2}{(\omega_{mk} + \omega)} + e^{i(\omega_{mk}-\omega)t/2}\frac{\sin(\omega_{mk} - \omega)t/2}{(\omega_{mk} - \omega)}\right] \qquad (3\ 3\text{-}16)$$

It can be easily seen that for large values of t, the function

$$\frac{\sin(\omega_{mk} - \omega)t/2}{\omega_{mk} - \omega} \qquad (3\ 3\text{-}17)$$

is very sharply peaked around $\omega \simeq \omega_{mk}$ and negligible everywhere else (see Fig 3 4) Thus, for states for which ω_{mk} is significantly different from ω, $C_m(t)$ would be negligible and transitions between such states will not be stimulated by the incident field This justifies our earlier statement that

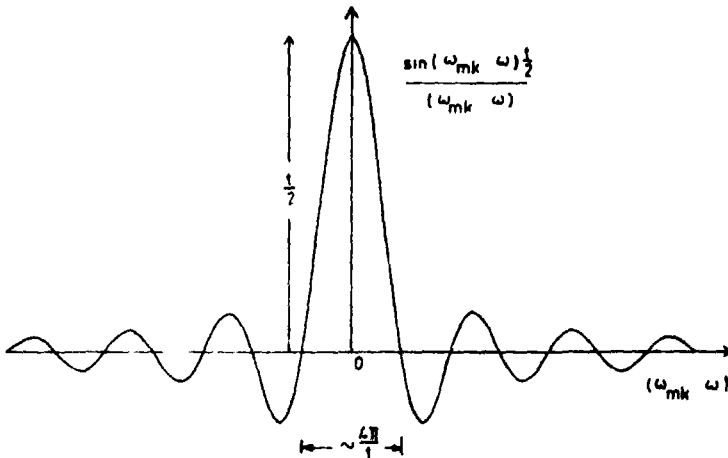



Fig 3 4 For large values of t the function given by Eq (3 3-17) is a very sharply peaked function of ω about $\omega = \omega_{mk}$

the presence of only those excited states be considered which are close to the resonance frequency

In an emission process, $\omega_k > \omega_m$ and hence ω_{mk} is negative thus it is the first term on the right-hand side of Eq (3 3-16) which contributes On the other hand in an absorption process, $\omega_{mk} > 0$ and the second term in Eq (3 3-16) contributes

We consider absorption of radiation and assume that at $t = 0$ the atom is in state 1 the corresponding wave function being $\psi_1(\mathbf{r})$ We also assume ω to be close to ω_{21} $[=(E_2 - E_1)/\hbar]$—see Fig 3 1 The probability for the transition to occur to state 2 is given by

$$|C_2(t)|^2 = \frac{1}{4}\frac{\mathscr{D}_{21}^2 E_0^2}{\hbar^2}\left\{\frac{\sin[(\omega_{21} - \omega)/2]t}{(\omega_{21} - \omega)/2}\right\}^2 \tag{3 3-18}$$

The above expression represents the probability for stimulated absorption of radiation In deriving Eq (3 3-18) we have assumed that $|C_2(t)|^2 \ll 1$, thus the result is accurate when

$$\left(\frac{\mathscr{D}_{21}E_0 t}{\hbar}\right)^2 \ll 1 \quad \text{or} \quad \left(\frac{\mathscr{D}_{21}E_0}{\hbar}\right)^2 \Big/ (\omega_{21} - \omega)^2 \ll 1 \tag{3 3-19}$$

A more exact result for a two-state system will be discussed in Section 3 4

We next assume that the quantity $(\omega_{21} - \omega)$ has a range of values either on account of the field having a continuous spectrum or the atom being capable of interacting with radiation having a range of frequencies

3 3 1 *Interaction with Radiation Having a Broad Spectrum*

We first consider the field having a continuous spectrum characterized by $u(\omega)$ which is defined such that $u(\omega)\,d\omega$ represents the energy associated with the field per unit volume within the frequency interval ω and $\omega + d\omega$ We replace† E_0^2 in Eq (3 3-18) by $(2/\epsilon_0)u(\omega)\,d\omega$ and integrate over all frequencies, which gives us the following expression for the transition probability

$$\Gamma_{12} = \frac{1}{2\epsilon_0}\frac{\mathscr{D}_{21}^2}{\hbar^2}\int u(\omega)\left\{\frac{\sin[(\omega_{21} - \omega)/2]t}{(\omega_{21} - \omega)/2}\right\}^2 d\omega \tag{3 3-20}$$

Assuming that $u(\omega)$ varies very slowly in comparison to the quantity inside the square brackets, we replace $u(\omega)$ by its value at $\omega = \omega_{21}$ and

† This follows from the fact that the average energy density associated with an electromagnetic wave is $\frac{1}{2}\epsilon_0 E_0^2$ where ϵ_0 is the dielectric permittivity of free space

take it out of the integral to obtain

$$\Gamma_{12} \simeq \frac{1}{2\epsilon_0}\frac{\mathscr{D}_{21}^2}{\hbar^2}u(\omega_{21})\left(\int_{-\infty}^{+\infty}\frac{\sin^2\xi}{\xi^2}\,d\xi\right)2t, \quad \xi = \frac{\omega_{21}-\omega}{2}t$$

$$= \frac{\pi}{\epsilon_0}\frac{\mathscr{D}_{21}^2}{\hbar^2}u(\omega_{21})t \qquad (3.3\text{-}21)$$

The above expression shows that† the probability of transition is proportional to t, thus the transition probability per unit time (which we denote by w_{12}) would be given by

$$w_{12} \simeq \frac{\pi}{\epsilon_0}\frac{\mathscr{D}_{21}^2}{\hbar^2}u(\omega_{21}) \qquad (3.3\text{-}22)$$

Now (omitting the subscripts) we have

$$\mathscr{D} = \hat{\mathbf{e}}\cdot\mathbf{P} = P\cos\theta \qquad (3.3\text{-}23)$$

where θ is the angle that $\hat{\mathbf{e}}$ (i e , the electric field) makes with the dipole moment vector $\mathbf{P}$ Assuming that the dipole moment vector is randomly oriented, the average value of $\mathscr{D}^2$ is given by

$$\overline{\mathscr{D}^2} = P^2\,\overline{\cos^2\theta} = \tfrac{1}{3}P^2 \qquad (3.3\text{-}24)$$

where use has been made of the following relation

$$\overline{\cos^2\theta} = \frac{1}{4\pi}\int_0^{2\pi}\int_0^{\pi}\cos^2\theta\sin\theta\,d\theta\,d\varphi = \tfrac{1}{3} \qquad (3.3\text{-}25)$$

Thus

$$w_{12} = \frac{\pi}{3\epsilon_0}\frac{P^2}{\hbar^2}u(\omega_{21}) \qquad (3.3\text{-}26)$$

If there are N_1 atoms per unit volume in the state 1 then the total number of absorptions per unit time per unit volume would be $N_1 w_{12}$, which would be equal to

$$N_1\frac{\pi}{3\epsilon_0}\frac{P^2}{\hbar^2}u(\omega_{21}) \qquad (3.3\text{-}27)$$

† It may be noted that Eq (3 3-21) predicts an indefinite increase in the transition probability with time however, the first-order perturbation theory itself breaks down when Γ_{12} is not appreciably less than unity Thus Eq (3.3-21) gives correct results as long as $\Gamma_{12} \ll 1$

Comparing Eqs (3 3-27) and (3 2-2), we obtain†

$$B_{12} = \frac{\pi}{3\epsilon_0}\frac{P^2}{\hbar^2} = \frac{4\pi^2}{3\hbar^2}\left(\frac{e^2}{4\pi\epsilon_0}\right)\left|\int \psi_2^* \mathbf{r} \psi_1 \, d\tau\right|^2 \tag{3 3-28}$$

The corresponding expression for stimulated emission is obtained by starting with the first term on the right-hand side of Eq (3 3-16) and proceeding in a similar manner The final expression is identical to Eq (3 3-28) except for an interchange of indices 1 and 2

Using Eq (3 2-10), we get the following expression for the A coefficient

$$A = \frac{4}{3}\left(\frac{e^2}{4\pi\epsilon_0}\frac{1}{\hbar c}\right)\frac{\omega^3}{c^2}\left|\int \psi_2^* \mathbf{r} \psi_1 \, d\tau\right|^2 \tag{3 3-29}$$

It may be of interest to note that

$$\frac{e^2}{4\pi\epsilon_0}\frac{1}{\hbar c} \simeq \frac{1}{137} \tag{3 3-30}$$

Using this value, we obtain

$$A \simeq \frac{4}{3}\frac{1}{137}\frac{\omega^3}{c^2}\left|\int \psi_2^* \mathbf{r} \psi_1 \, d\tau\right|^2 \tag{3 3-31}$$

As an example, we calculate the A coefficient for the $2P \rightarrow 1S$ transition in the hydrogen atom, i e , the transition from the ($n = 2$, $l = 1$, $m = 0$) state to the ($n = 1$, $l = 0$ $m = 0$) state For these states (see, e g , Ghatak and Lokanathan, 1977)

$$\psi_1 = \frac{1}{(4\pi)^{1/2}}\frac{2}{a_0^{3/2}}\exp\left(-\frac{r}{a_0}\right) \tag{3 3-32}$$

$$\psi_2 = \frac{1}{(2a_0)^{3/2}}\frac{r}{a_0 3^{1/2}}\exp\left(-\frac{r}{2a_0}\right)\left[\left(\frac{3}{4\pi}\right)^{1/2}\cos\theta\right] \tag{3 3-33}$$

where $a_0 = (\hbar^2/m)(4\pi\epsilon_0/e^2) \simeq 0.5 \times 10^{-10}$ m In order to evaluate the matrix element, we write

$$\begin{aligned} x &= r \sin\theta \cos\varphi \\ y &= r \sin\theta \sin\varphi \\ z &= r \cos\theta \end{aligned} \tag{3 3-34}$$

† The corresponding expression for B_{12} in the CGS units is obtained by replacing $(e^2/4\pi\epsilon_0)$ by e^2 where e (the charge of the electron) is now measured in the CGS system of units

Now,

$$\int \psi_1^* x \psi_2 \, d\tau = \frac{1}{4\pi(2)^{1/2}} \frac{1}{a_0^4} \left(\int_0^\infty r^2 \, dr \, e^{-3r/2a_0} r^2 \right) \times \left(\int_0^\pi \cos\theta \sin^2\theta \, d\theta \right) \left(\int_0^{2\pi} \cos\varphi \, d\varphi \right) = 0 \tag{3 3-35}$$

because the integral over φ vanishes Similarly

$$\int \psi_1^* y \psi_2 \, d\tau = 0 \tag{3 3-36}$$

The only nonvanishing integral is

$$\int \psi_1^* z \psi_2 \, d\tau = \left(\frac{1}{4\pi(2)^{1/2}} \frac{1}{a_0^4} \right) \int_0^\infty r^2 \, dr \, e^{-3r/2a_0} r^2 \int_0^\pi \cos^2\theta \sin\theta \, d\theta \int_0^{2\pi} d\varphi = 4(2^{1/2})\left(\tfrac{2}{3}\right)^5 a_0 \tag{3 3-37}$$

Thus†

$$\left| \int \psi_1^* \mathbf{r} \psi_2 \, d\tau \right|^2 = 2^5 \left(\tfrac{2}{3}\right)^{10} a_0^2 \tag{3.3-38}$$

Further, for the $2P \rightarrow 1S$ transition

$$\omega = \frac{1}{\hbar} \frac{3}{8a_0} \left(\frac{e^2}{4\pi\epsilon_0} \right) = \frac{3c}{8a_0} \left(\frac{e^2}{4\pi\epsilon_0} \frac{1}{\hbar c} \right) \simeq \frac{3 \times 3 \times 10^8}{8 \times 0\ 51 \times 10^{-10}} \frac{1}{137} \simeq 1\ 5 \times 10^{16} \text{ sec}^{-1} \tag{3 3-39}$$

Substituting in Eq (3 3-31), we obtain

$$A \simeq \frac{4}{3} \ \frac{1}{137} \ \frac{[1\ 5 \times 10^{16}]^3}{(3 \times 10^8)^2} 2^5 \left(\tfrac{2}{3}\right)^{10} (0\ 5 \times 10^{-10})^2 \simeq 6 \times 10^8 \text{ sec}^{-1} \tag{3 3-40}$$

The mean lifetime of the state, τ, is the inverse of A giving

$$\tau \simeq 1\ 6 \times 10^{-9} \text{ sec} \tag{3 3-41}$$

† It can be shown that $|\int \psi_1^* \mathbf{r} \psi_2 \, d\tau|^2$ has the same value for transition from any one of the states ($n = 2$ $l = 1$, $m = 0$) or ($n = 2$, $l = 1$, $m = -1$) or ($n = 2$ $l = 1$ $m = -1$) to the ($n = 1$ $l = 0$ $m = 0$) state However, the matrix element for the transition from ($n = 2$, $l = 0$ $m = 0$) state to the ($n = 1$ $l = 0$ $m = 0$) state is zero This implies that the corresponding dipole transition is forbidden

Thus the lifetime of the hydrogen atom in the upper level corresponding to the $2P \rightarrow 1S$ transition is about 1.6×10^{-9} sec. Transitions having such small lifetimes are referred to as strongly allowed transitions.

In contrast, the levels used in the laser transition are such that the upper level has a very long lifetime ($\sim 10^{-3}$–10^{-5} sec). A level having such a long lifetime is referred to as a metastable level, and such transitions come under the class of weakly allowed or nearly forbidden transitions. The strength of an atomic transition is usually expressed in terms of the f value defined by the following equation:

$$f_{21} = \frac{2}{3}\frac{m\omega_{21}}{\hbar}|\mathscr{D}_{21}|^2 \tag{3.3-42}$$

For strongly allowed transitions f is of the order of unity; for example, for the $2P \rightarrow 1S$ transition in the hydrogen atom, $f = 0.416$. On the other hand, for the transitions from the upper laser level $f \sim 10^{-3}$–10^{-6}.

3.3.2 Interaction of a Near-Monochromatic Wave with an Atom Having a Broad Frequency Response

We next consider a nearly monochromatic field [see Eq. (3.2-15)] interacting with atoms characterized by the line-shape function $g(\omega)$. For such a case the probability for the atom being in the upper state would be given by

$$\begin{aligned}\Gamma_{12} &= \frac{1}{4}\frac{\mathscr{D}_{21}^2 E_0^2}{\hbar^2}\int g(\omega')\left[\frac{\sin(\omega'-\omega)t/2}{(\omega'-\omega)/2}\right]^2 d\omega' \\ &\simeq \frac{1}{4}\frac{\mathscr{D}_{21}^2}{\hbar^2}E_0^2 g(\omega)2\pi t \\ &\simeq \frac{\pi P^2}{3\hbar^2\epsilon_0}g(\omega)u_\omega t\end{aligned} \tag{3.3-43}$$

where in the last step we have replaced $\mathscr{D}_{21}^2$ and E_0^2 by $\frac{1}{3}P^2$ [see Eq. (3.3-24)] and $2u_\omega/\epsilon_0$ (see footnote on p. 46), respectively. Since [see Eq. (3.2-12)]

$$B_{12} = B_{21} = \frac{\pi}{3\epsilon_0}\frac{P^2}{\hbar^2} = \frac{\pi^2 c^3}{\hbar\omega^3 t_{sp}} \tag{3.3-44}$$

we obtain the following expression for the transition rate (per unit time) per unit volume:

$$W_{12} = N_1\frac{\pi^2 c^3}{\hbar\omega^3 t_{sp}}u_\omega g(\omega) \tag{3.3-45}$$

which is consistent with Eq. (3.2-17).

3 4. More Accurate Solution for the Two-Level System

A more accurate solution of the time-dependent Schrodinger equation [see Eqs (3 3-2) and (3 3-3)] can be obtained if we assume that the atom can exist in only two possible states characterized by $\psi_1(\mathbf{r})$ and $\psi_2(\mathbf{r})$ Thus, Eq (3 3-8) gets replaced by

$$\Psi(\mathbf{r}, t) = C_1(t)\psi_1(\mathbf{r})e^{-i\omega_1 t} + C_2(t)\psi_2(\mathbf{r})e^{-i\omega_2 t} \tag{3 4-1}$$

If we substitute from Eq (3 4-1) into Eq (3 3-2), multiply by ψ_1^*, and integrate, we would get [cf Eq (3 3-15)]

$$i\hbar \frac{dC_1}{dt} = \tfrac{1}{2}E_0\mathscr{D}_{12}C_2(t)(e^{-i(\omega - \omega')t} + e^{-i(\omega + \omega')t}) \tag{3 4-2}$$

Similarly,

$$i\hbar \frac{dC_2}{dt} = \tfrac{1}{2}E_0\mathscr{D}_{21}C_1(t)(e^{i(\omega + \omega')t} + e^{i(\omega - \omega')t}) \tag{3 4-3}$$

where use has been made of the fact that

$$\int \psi_1^*\mathbf{r}\psi_1 \, d\tau = \int \psi_2^*\mathbf{r}\psi_2 \, d\tau = 0 \tag{3 4-4}$$

and

$$\omega' = (E_2 - E_1)/\hbar = \omega_{21} = -\omega_{12}$$

In the rotating wave approximation, considering absorption we neglect the terms $e^{-i(\omega + \omega')t}$ and $e^{+i(\omega + \omega')t}$ in Eqs (3 4-2) and (3 4-3) and obtain†

$$\frac{dC_1}{dt} = -\frac{i}{2\hbar} E_0\mathscr{D}_{12}C_2(t)e^{i(\omega - \omega')t} \tag{3 4-5}$$

$$\frac{dC_2}{dt} = -\frac{i}{2\hbar} E_0\mathscr{D}_{21}C_1(t)e^{-i(\omega - \omega')t} \tag{3 4-6}$$

If we assume a solution of the form

$$C_1(t) = e^{i\Omega t} \tag{3 4-7}$$

† This can be understood if we refer to Eq (3 3-16) where the first term on the right-hand side can be neglected with respect to the second term since $\omega \approx \omega'$ Salzman (1971) has obtained an exact numerical solution of the Schrödinger equation for a two-state system without using the rotating wave approximation The final result for $|C_2(t)|^2$ is the same except for the presence of a weak high-frequency oscillation superimposed on the solution

then from Eq (3.4-5),

$$C_2(t) = -\frac{2\hbar\Omega}{E_0\mathscr{D}_{12}} e^{i(\Omega-\omega+\omega')t} \tag{3.4-8}$$

Substituting this in Eq (3.4-6), we get

$$-i\frac{2\hbar\Omega}{E_0\mathscr{D}_{12}}(\Omega - \omega + \omega') = -\frac{i}{2\hbar}E_0\mathscr{D}_{21}$$

or

$$\Omega(\Omega + \omega' - \omega) - \frac{\Omega_0^2}{4} = 0 \tag{3.4-9}$$

where

$$\Omega_0^2 = \frac{\mathscr{D}_{12}\mathscr{D}_{21}E_0^2}{\hbar^2} = \frac{\mathscr{D}^2 E_0^2}{\hbar^2} \tag{3.4-10}$$

and

$$\mathscr{D} = \mathscr{D}_{12} = \mathscr{D}_{21}$$

Equation (3.4-9) gives

$$\Omega_{1,2} = \tfrac{1}{2}\{-(\omega' - \omega) \pm [(\omega' - \omega)^2 + \Omega_0^2]^{1/2}\} \tag{3.4-11}$$

Thus, the general solution will be

$$C_1(t) = A_1 e^{i\Omega_1 t} + A_2 e^{i\Omega_2 t} \tag{3.4-12}$$

$$C_2(t) = -\frac{2}{\Omega_0} e^{i(\omega-\omega')t}(A_1\Omega_1 e^{i\Omega_1 t} + A_2\Omega_2 e^{i\Omega_2 t}) \tag{3.4-13}$$

If we now assume that the atom is initially in the ground state, i.e.

$$C_1(0) = 1, \qquad C_2(0) = 0 \tag{3.4-14}$$

then

$$A_1 = -\frac{\Omega_2}{\Omega_1} A_2 \tag{3.4-15}$$

and

$$1 = A_1 + A_2 = A_2\frac{\Omega_1 - \Omega_2}{\Omega_1} = [(\omega' - \omega)^2 + \Omega_0^2]^{1/2}\ \frac{A}{\Omega_1}$$

or

$$A_2 = \frac{\Omega_1}{\Omega'} \tag{3.4-16}$$

where

$$\Omega' = [(\omega' - \omega)^2 + \Omega_0^2]^{1/2} \tag{3.4-17}$$

On substitution we finally obtain

$$C_2(t) = -i\frac{\Omega_0}{\Omega'} e^{i(\omega' - \omega)t/2} \sin \Omega' t/2 \tag{3.4-18}$$

Thus the transition probability for absorption is given by

$$|C_2(t)|^2 = \left(\frac{\sin \Omega' t/2}{\Omega'/2}\right)^2 \left(\frac{\Omega_0}{2}\right)^2 \tag{3.4-19}$$

which has been plotted in Fig 3.5 Also shown in the figure are the results of the exact numerical calculations without resorting to the rotating wave approximation

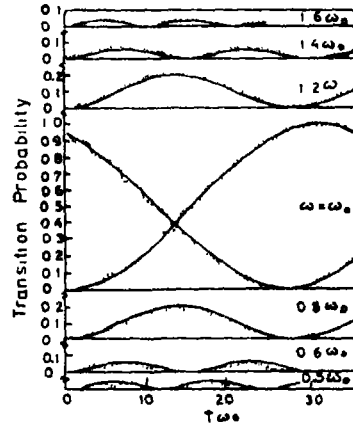



Fig 3.5 Variation of the transition probability as a function of time for a two level system for different frequencies ω of the electromagnetic field The curves correspond to $\mathscr{D}E_0/\hbar = 0.1\omega$ The solid line corresponds to Eq (3.4-19) and the dotted curve corresponds to an exact numerical computation Notice the presence of a weak high-frequency oscillation present in the exact calculation (After Salzman 1971)

At resonance, $\omega = \omega'$ and one obtains

$$|C_2(t)|^2 = \sin^2 \frac{\Omega_0 t}{2} \tag{3.4-20}$$

which shows that the system flip-flops between states 1 and 2 A comparison of Eqs (3.4-19) and (3.3-18) shows that the perturbation theory result is valid if

$$\left(\frac{\mathscr{D}_{21}E_0}{\hbar}t\right)^2 \ll 1 \quad \text{or} \quad \left(\frac{\mathscr{D}_{21}E_0}{\hbar}\right)^2 \frac{1}{(\omega' - \omega)^2} \ll 1 \tag{3.4-21}$$

It may be of interest to note that the solutions obtained in this section are exact when $\omega = 0$ (i e , a constant electric field) and if $\mathscr{D}_{12}$ is replaced by $2\mathscr{D}_{12}$ in the solution given by Eq (3 4-19) This follows from the fact that for $\omega = 0$, the exact equations [Eqs (3 4-2) and (3 4-3)] are the same as Eqs (3 4-5) and (3 4-6) with $\mathscr{D}_{12}$ replaced by $2\mathscr{D}_{12}$

3 5 Line-Broadening Mechanisms

In this section we will discuss some important line-broadening mechanisms and the corresponding line-shape functions A study of the line-broadening mechanism (and hence the line-shape function) is of great importance as it determines the operation characteristics of the laser e g , the threshold population inversion [see Eq (3 2-29)], the number of oscillating modes (see Chapter 6) etc

We first consider the line-shape function corresponding to collisions that occur in a collection of atoms In a collection of atoms in the gaseous form random collisions occur and thus an atom when interacting with an electromagnetic field sees a field which changes its phase abruptly at each collision Thus if the average time between two collisions is τ then the atom no longer sees a monochromatic wave but instead a wave like that shown in Fig 3 6, where the field makes random changes in phase at every collision Thus the line-shape function would be given by (apart from some proportionality constant) the power spectrum of the field shown in Fig 3 6 In order to evaluate this we note that the field of the type shown in Fig 3 6 can be written in the form

$$E(t) = E_0 \exp(-i\omega_0 t + i\phi) \tag{3 5-1}$$

where the phase ϕ remains constant for

$$t_0 \leqslant t \leqslant t_0 + \tau \tag{3 5-2}$$

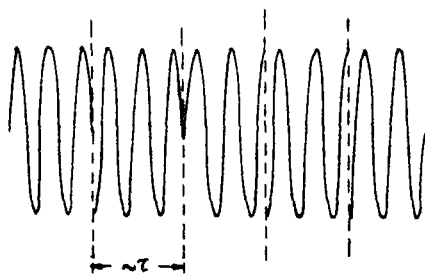



Fig 3 6 When an atom undergoes random collisions it sees a wave like that shown in the figure At every collision the phase of the wave changes abruptly

and at each collision the phase ϕ changes randomly The frequency spread of such a wave would be obtained by carrying out a Fourier transform

$$E(\omega) = \frac{1}{2\pi}\int_{t_0}^{t_0+\tau} E_0 e^{-i\omega_0 t + i\phi} e^{i\omega t}\, dt$$
$$= \frac{1}{2\pi} E_0 \exp[i(\omega - \omega_0)t_0 + i\phi] \frac{\exp[i(\omega - \omega_0)\tau] - 1}{i(\omega - \omega_0)}$$

Thus the frequency distribution of the intensity would be given by

$$I(\omega) \propto |E(\omega)|^2 = \left(\frac{E_0}{\pi}\right)^2 \frac{\sin^2[(\omega - \omega_0)\tau/2]}{(\omega - \omega_0)^2} \qquad (3\ 5\text{-}3)$$

Now at any instant the radiation is from atoms that have different values of τ thus, in order to obtain the spectral density we must multiply Eq (3 5-3) by $P(\tau)\, d\tau$ and integrate from 0 to ∞, here $P(\tau)\, d\tau$ represents the probability that the atom suffers a collision after a time interval between τ and $\tau + d\tau$ It can be shown† from the kinetic theory that

$$P(\tau)\, d\tau = \frac{1}{\tau_0} e^{-\tau/\tau_0}\, d\tau \qquad (3\ 5\text{-}4)$$

where τ_0 represents the mean time between two collisions Notice that

$$\int_0^\infty P(\tau)\, d\tau = 1 \qquad (3\ 5\text{-}5)$$

and

$$\int_0^\infty \tau P(\tau)\, d\tau = \tau_0 \qquad (3\ 5\text{-}6)$$

Thus if $g(\omega)$ represents the frequency distribution of the radiation causing the transition then

$$g(\omega) \propto \int_0^\infty \frac{\sin^2[(\omega - \omega_0)\tau/2]}{(\omega - \omega_0)^2} e^{-\tau/\tau_0} \frac{1}{\tau_0}\, d\tau$$
$$= \tfrac{1}{2}[(\omega_0 - \omega)^2 + (1/\tau_0)^2]^{-1}$$

or

$$g(\omega) = \frac{\tau_0}{\pi} \frac{1}{1 + (\omega_0 - \omega)^2 \tau_0^2} \qquad (3\ 5\text{-}7)$$

† See e g, Gopal (1974) Chap 4

where the constant factor has been adjusted such that

$$\int g(\omega)\, d\omega = 1 \tag{3 5-8}$$

The distribution $g(\omega)$ given by Eq (3 5-7) is known as a Lorentzian and is plotted in Fig 3 7 The peak of $g(\omega)$ lies at $\omega = \omega_0$ and has a value τ_0/π The full width at half maximum is given by

$$\Delta\omega = 2/\tau_0$$

In a similar manner, one can calculate the effect of the thermal motions of the gas atoms In a gas the atoms undergo random motions When such a moving atom interacts with radiation, the apparent frequency of the incident wave is different from the frequency seen from a

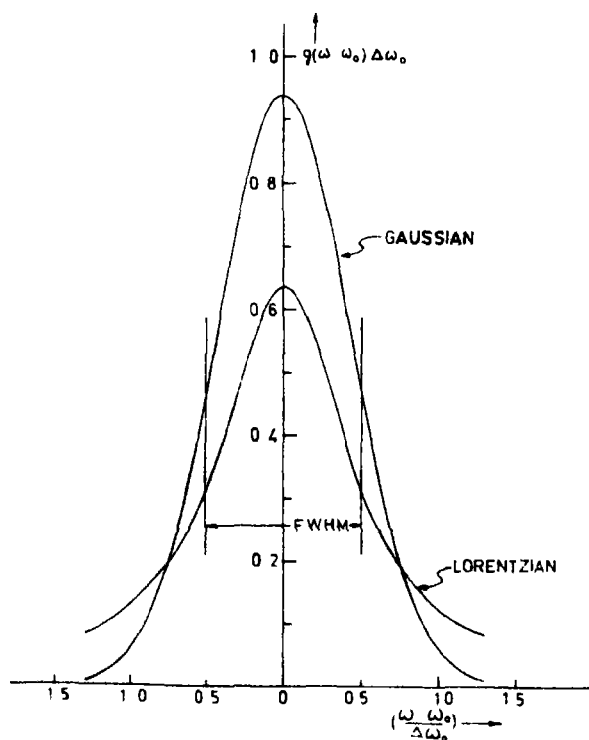



Fig 3 7 Lorentzian and Gaussian line shapes [see Eqs (3 5 19) and (3 5 15)] corresponding to the same width at half maximum Notice that the Gaussian has a larger peak value for a given half width and is also much sharper as compared with a Lorentzian curve

stationary atom, this is called the Doppler effect. Equivalently, one may say that the resonance frequency of the atom is shifted due to the motion Thus if ω represents the frequency of the incident wave, then an atom moving with a velocity whose component in the direction of the radiation is v_z sees a frequency

$$\tilde{\omega} = \omega\left(1 - \frac{v_z}{c}\right) \tag{3 5-9}$$

Thus, if ω_{21} is the transition frequency, in order that the incident radiation may interact strongly with the atom, its frequency must be

$$\omega = \omega_{21}\left(1 - \frac{v_z}{c}\right)^{-1} \simeq \omega_{21}\left(1 + \frac{v_z}{c}\right) \tag{3 5-10}$$

Thus the effect of the motion is to change the resonant frequency of the atom

In order to calculate the line-shape function, we note that the probability that an atom has a z component of velocity lying between v_z and $v_z + dv_z$ is given by the Maxwell distribution

$$P(v_z)\,dv_z = \left(\frac{M}{2\pi k_B T}\right)^{1/2} \exp\left(-\frac{Mv_z^2}{2k_B T}\right) dv_z \tag{3 5-11}$$

where M is the mass of the atom and T the temperature of the gas Thus the probability $g(\omega)\,d\omega$ that the transition frequency lies between ω and $\omega + d\omega$ is equal to the probability that the z component of the velocity of the atom lies between v_z and $v_z + dv_z$, where

$$v_z = \frac{(\omega - \omega_{21})c}{\omega_{21}} \tag{3 5-12}$$

(We are assuming that the incident radiation is propagating along the z direction) Thus

$$g(\omega)\,d\omega = \frac{c}{\omega_{21}}\left(\frac{M}{2\pi k_B T}\right)^{1/2} \exp\left[-\frac{Mc^2}{2k_B T}\frac{(\omega - \omega_{21})^2}{\omega_{21}^2}\right] d\omega \tag{3 5-13}$$

This corresponds to a Gaussian distribution and is plotted in Fig 3 7 The distribution given by Eq (3 5-13) is peaked at $\omega = \omega_{21}$ and has a maximum value of $(2/\pi^{1/2})[(\ln 2)^{1/2}/\Delta\omega]$ at $\omega = \omega_{21}$, where $\Delta\omega$ is the full width at half maximum (FWHM) given by

$$\Delta\omega = 2\omega_{21}\left(\frac{2k_B T}{Mc^2}\ln 2\right)^{1/2} \tag{3 5-14}$$

Thus, in terms of $\Delta\omega$, the Gaussian line-shape function can be written as

$$g(\omega) = \frac{2}{\Delta\omega}\left(\frac{\ln 2}{\pi}\right)^{1/2} \exp\left[-4\ln 2\frac{(\omega - \omega_{21})^2}{(\Delta\omega)^2}\right] \qquad (3\ 5\text{-}15)$$

Even in the absence of collisions and of thermal motions, one has the inherent linewidth because of the finite lifetime of the excited state due to spontaneous emission The electric field associated with the spontaneous emission can be assumed to be of the form

$$E = E_0 \exp(-t/2t_{sp})e^{-i\omega_0 t} \qquad (3\ 5\text{-}16)$$

where t_{sp} is the spontaneous emission lifetime The frequency spectrum corresponding to the field given by Eq (3 5-16) is

$$\tilde{E}(\omega) = \int_0^\infty E_0 e^{-i\omega_0 t} e^{-t/t_{sp}} e^{i\omega t}\, dt$$

$$= E_0 \frac{1}{-i(\omega - \omega_0) + 1/2t_{sp}} \qquad (3\ 5\text{-}17)$$

Thus, the frequency spectrum of the intensity distribution will be given by

$$I(\omega) \propto |\tilde{E}(\omega)|^2 = \frac{E_0^2}{(\omega - \omega_0)^2 + (1/2t_{sp})^2} \qquad (3\ 5\text{-}18)$$

which is again a Lorentzian with a spectral distribution given by

$$g(\omega) = \frac{2t_{sp}}{\pi}\frac{1}{1 + 4(\omega_0 - \omega)^2 t_{sp}^2} = \frac{2}{\pi\,\Delta\omega_0}\frac{1}{1 + [(\omega_0 - \omega)/\frac{1}{2}\Delta\omega_0]^2} \qquad (3\ 5\text{-}19)$$

and a full width at half maximum given by $\Delta\omega_0 = 1/t_{sp}$ This kind of broadening is referred to as natural broadening [see also Eq (5 3-45) and Appendix F]

In the above we have considered the broadening produced by different mechanisms separately In general, all the mechanisms may be present simultaneously and the resultant line-shape function can be obtained by performing a convolution of the different line shapes If any one of the broadening mechanisms dominates over the others, then the line-shape function would correspond to the dominant mechanism Thus, for example, if we consider the 6328-Å transition of Ne in the He–Ne laser (see Section 9 4), at $T = 300°$K, the Doppler-broadened linewidth is given by [see Eq (3 5-14)]

$$\Delta\nu = \frac{\Delta\omega}{2\pi} = 2\nu_{12}\left(\frac{2k_BT}{Mc^2}\ln 2\right)^{1/2}$$

$$\simeq 1\,7 \times 10^9\ \text{sec}^{-1} = 1700\ \text{MHz} \qquad (3\ 5\text{-}20)$$

r typical He–Ne laser parameters, the collision broadening produces a ;width of about 0.64 MHz at a pressure of 0 5 Torr, while the natural ıadenıng ıs about 20 MHz Thus, for He–Ne laser parameters, the ppler broadenıng domınates over natural broadenıng and collision ıadenıng

The various lıne-broadenıng mechanısms can be broadly classified ıer homogeneous and inhomogeneous broadening Certain lıne-ıadenıng mechanısms lıke collısıon broadenıng or natural broadenıng to broaden the response of each atom ın an ıdentıcal fashıon, and such ıadenıng mechanısms come under the class of homogeneous broaden- On the other hand, Doppler broadenıng or broadenıng produced by al inhomogeneities in a crystal lattıce act to shıft the central frequency the response of ındıvıdual atoms by dıfferent amounts, thereby leadıng a broadened response Such a form of broadenıng ıs referred to as omogeneous broadenıng If the effects whıch cause the ınhomo-ıeous broadenıng are random in origin, then the broadened line is ussian in shape In contrast homogeneous broadenıng ıs, ın general, a rentzıan

6 Saturatıon Behavıor of Homogeneously and Inhomogeneously Broadened Transıtıons

In Sectıon 3 5 we dıscussed the varıous lıne-broadenıng mechanısms ongıng to both homogeneous and ınhomogeneous broadenıngs In thıs tıon, we brıefly dıscuss the dıfference ın saturatıon behavıor between two kınds of broadenıngs Let us first consıder a homogeneously ıadened laser medıum placed inside a resonator and let us also assume t there ıs a resonator mode coıncıdıng exactly wıth the center of the ' Inıtıally as the pumpıng rate ıs below threshold, the gaın ın the onator ıs less than the losses and the laser does not oscıllate As the npıng rate ıs ıncreased, first to reach threshold ıs the mode at the lıne ıter as ıt has the mınımum threshold We have seen earlıer that when laser ıs oscıllatıng ın steady state, the gaın ıs exactly equal to the ıes at the oscıllatıng frequency Thus, at steady state, even when the npıng power ıs ıncreased beyond threshold, the gaın at the oscıllatıng quency does not ıncrease beyond the threshold value, thıs ıs because of fact that the losses remaın constant In fact, the ıncrease ın pumpıng ver wıll be accompanıed by an ıncrease ın the power ın the mode ıch ın turn would be accompanıed by a stronger saturatıon of the laser ısıtıon, thus reducıng the gaın agaın to the value at threshold It may

be mentioned that the gain could exceed the threshold value on a transient basis but not under steady state operation

Now, in a homogeneously broadened transition, all the atoms have identical line shapes peaked at the same frequency Thus all atoms interact with the same oscillating mode and the increase in pumping power cannot increase the gain at other frequencies and thus the laser will oscillate only in a single mode This observation has indeed been verified experimentally on some homogeneously broadened transitions such as in a Nd YAG laser The fact that a laser with homogeneously broadened line shape can oscillate in many modes is due to spatial hole burning (see Section 5 3) This can be understood from the fact that each mode is a standing wave pattern between the resonator mirrors Thus there are regions of high population inversion (where the nodes of the fields exist) and regions of saturated population inversion (at antinodes of the field where the amplitude of the field is maximum) If one considers another mode which has (at least over some portions) antinodes at the nodes corresponding to the central oscillating mode, then this mode can draw energy from the atoms and, if the loss can be compensated by gain, this mode can also oscillate

In contrast to the case of homogeneous broadening, if the laser medium is inhomogeneously broadened then a given mode at a certain frequency can interact with only a group of atoms whose response curve contains the mode frequency (see Fig 3 8) Thus if the pumping is increased beyond threshold, the gain at the oscillating mode frequency remains fixed but the gain at other frequencies can go on increasing (see Fig 3 9) Thus, in an inhomogeneously broadened line one can have multimode oscillation and as one can see from Fig 3 9 each oscillating

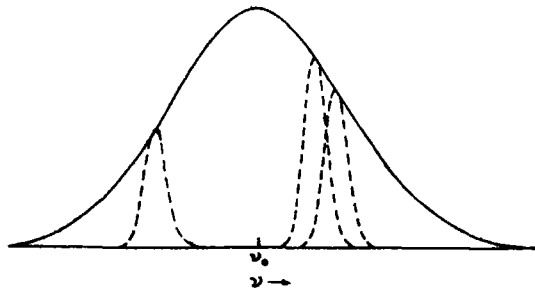



Fig 3 8 The inhomogeneous line is formed by a group of atoms which have slightly different resonance frequencies on the same transition An applied signal at a certain frequency will interact strongly only with those atoms whose resonance frequencies are within about a natural linewidth from the applied frequency Thus the applied signal interacts only with a group of atoms

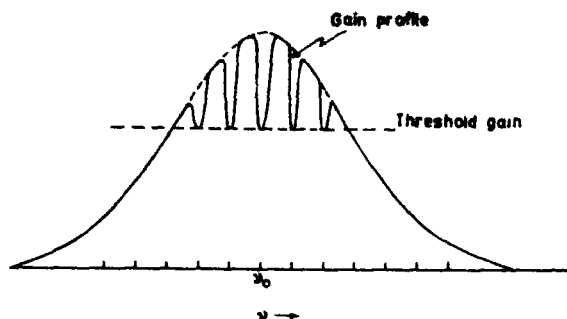



Fig 3 9 As the pumping is increased beyond threshold under steady state operation the gain at the various oscillating frequencies cannot increase beyond the threshold value but the gain at other frequencies may be much above the threshold value The various oscillating modes are said to burn holes in the gain curve The dashed line represents the gain in the absence of the resonator mirrors

mode "burns holes in the frequency space" of the gain profile These general conclusions regarding homogeneously and inhomogeneously broadened lines have indeed been verified experimentally

Various techniques for single longitudinal mode oscillation of inhomogeneously broadened lasers are discussed in Section 6 5

Let us now consider an inhomogeneously broadened laser medium and let us assume that only a single mode exists within the entire gain profile Let us also assume, to begin with, that the frequency of the mode does not coincide with the line center and that we slowly change the frequency of the mode so that it passes through the center of the profile to the other side of the peak in the gain profile† In order to determine the variation of the power output as the frequency is scanned through the line center, we observe that a mode of the laser is actually made up of two traveling waves traveling along opposite directions along the resonator axis Thus when the mode frequency does not coincide with the line center, the wave traveling from left to right in the resonator will interact with those atoms whose z-directed velocities are near to [see Eq (3 5-12)]

$$v_z = \frac{\omega - \omega_{21}}{\omega_{21}} c \qquad (3\ 6\text{-}1)$$

while the wave moving from right to left would interact with those atoms

† Such a scan could for example be performed by mounting one of the cavity mirrors on a piezoelectric crystal and changing the cavity length by an applied field

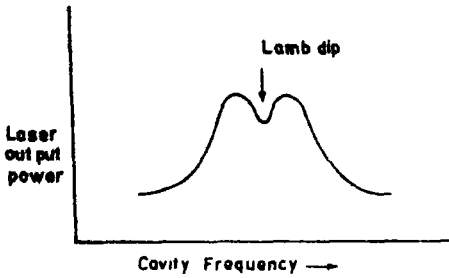



Fig 3 10 As the frequency of the oscillating cavity mode is scanned through the center of the line a dip in the power output is observed for inhomogeneous transitions This phenomenon is referred to as the Lamb dip

whose z-directed velocity would be

$$v_z = -\frac{\omega - \omega_{21}}{\omega_{21}} c \tag{3 6-2}$$

Thus there are two groups of atoms with equal and opposite z-directed velocities which are strongly interacting with the mode As the frequency of the mode is tuned to the center these group of atoms change with the frequency, and at the line center, the mode can interact only with the group of atoms having a zero value of z-directed velocity Thus the power output must decrease slightly when the mode frequency is tuned through the line center In fact, this has been observed experimentally and is referred to as the Lamb dip (see Fig 3 10)—the presence of a Lamb dip in a He–Ne laser was shown by McFarlane, Bennett, and Lamb (1963)

4

Laser Rate Equations

4.1. Introduction

In Chapter 3 we studied the interaction of radiation with matter and found that under the action of radiation of proper frequencies, the atomic populations of various energy levels change In the present chapter, we will be studying the rate equations which govern the rate at which populations of various levels change under the action of a pump and in the presence of the laser radiation The rate equation approach provides a convenient means of studying the time dependence of the atomic populations of various levels under the presence of radiation at frequencies corresponding to the different transitions of the atoms It also gives the steady state population difference between the actual levels involved in the laser transition and allows one to study whether an inversion of population is achievable in a transition and, if so, what would be the minimum pumping rate required to maintain a steady population inversion for continuous wave operation of the laser Once we have thus determined the conditions for obtaining population inversion between two levels, the gain that such a medium would provide at and near the transition frequency and the phase shift effects that such a medium would introduce are discussed in detail in Chapter 5 Thus Chapter 5 discusses the behavior of a system having two levels when there is a population inversion between the two levels, and the present chapter deals with the means of obtaining an inversion between the two levels of an atomic system by making use of other energy levels The rate equations can also be solved to obtain the transient behavior of the laser, which gives rise to phenomena like laser spiking

The atomic rate equations along with the rate equation for the photon number in the cavity form a set of nonlinear equations These equations can be solved under the steady state regime and one can study the evolution of the photon number as one passes through the threshold pumping region

In Sections 4 2 and 4 3, we consider in detail the three-level and the four-level laser systems We obtain the relations to be satisfied by the relaxation rates of the various levels in order that population inversion may be attained We show that in the three-level laser system, a minimum pump power is required to obtain an inversion and hence an amplifying medium In contrast, in an ideal four-level system, the attainment of inversion is independent of the pump power but the magnitude of population inversion does depend on the pumping rate

We also saw in Section 3 2 that in order that the laser act as an oscillator one requires a minimum inversion density Since the amount of inversion depends on the pump power, we see that there is a minimum pumping rate for obtaining oscillation in a laser cavity In Sections 4 2 and 4 3 we obtain this critical pump power as a function of the cavity parameters Typical examples of three-level and four-level systems are also given

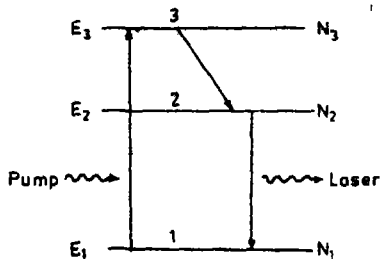
In Section 4 4 we study a simple two-level system and discuss the evolution of the number of photons in the cavity as one passes through the threshold region In Section 4 5 we obtain the optimum output coupling for obtaining maximum laser power, and in Section 4 6 we briefly discuss the phenomenon of laser spiking

4 2 The Three-Level System

Consider a three-level laser system and assume that all the levels are nondegenerate (see Fig 4 1) The pump is applied in the $1 \rightarrow 3$ transition and the lasing transition is $2 \rightarrow 1$ In this system the pump lifts atoms from the level 1 into the level 3, from which they decay rapidly to level 2 through some nonradiative† process, the level 2 is required to be metastable‡ Thus the pump effectively transfers atoms from level 1 to level 2 through level 3 If the relaxation from level 3 to level 2 is very fast, then most of the atoms in level 3 will relax down to level 2 rather than to level 1 Since the upper pump level 3 is not one of the laser levels, level 3 can be a broad level (or a group of broad levels) so that a broadband light source can be used efficiently as a pumping source Also, since the lower laser level is the ground level, more than 50% of the atoms from level 1 have to be lifted up to attain population inversion between levels 2 and 1

† In a nonradiative decay process, the energy is given to the surrounding molecules in the form of translational vibrational, or rotational energy In a solid, the energy is given to the lattice of atoms increasing the lattice energy

‡ See Section 3 3

Fig 4 1 The energy levels in a three-level laser system Level 1 is the ground level and levels 2 and 3 are excited levels The pump lifts atoms from level 1 to level 3 from which they decay rapidly to level 2 Population inversion is obtained between levels 2 and 1 and the laser oscillates at the frequency corresponding to the 2 → 1 transition

Let N_1, N_2, and N_3 represent the number of atoms per unit volume in levels 1, 2, and 3, respectively We assume that only these three levels are populated and that the transitions take place only between these three levels For such a case, we may write

$$N = N_1 + N_2 + N_3 \tag{4 2-1}$$

where N represents the total number of atoms per unit volume We now write down the rate at which the population of any level changes For example, the change in the population of level 3 is described by the following equation

$$\frac{dN_3}{dt} = W_p(N_1 - N_3) - N_3 T_{32} \tag{4 2-2}$$

where $W_p N_1$ represents the number of induced absorptions per unit time (per unit volume) which results in the $1 \rightarrow 3$ transition, similarly $W_p N_3$ represents the number of stimulated emissions per unit time (per unit volume) associated with the $3 \rightarrow 1$ transition† The last term in Eq (4 2-2) represents the $3 \rightarrow 2$ transition caused mainly by extremely fast nonradiative processes‡ Further,

$$T_{32} = A_{32} + S_{32} \tag{4 2-3}$$

where A_{32} represents the Einstein A coefficient (corresponding to a radiative transition) connecting levels 3 and 2, and S_{32} represents the nonradiative transition rate from level 3 to level 2

Similarly we can write for the rate of change of the population of level 2

$$\frac{dN_2}{dt} = W_l(N_1 - N_2) + N_3 T_{32} - N_2 T_{21} \tag{4 2-4}$$

† The quantity W_p is proportional to the Einstein coefficient B_{13} and to the pump beam energy density [see Eq (3 2-4)]

‡ We have neglected the contribution due to $3 \rightarrow 1$ spontaneous transitions, this is justified as in most practical laser systems, the atoms in level 3 almost instantaneously undergo a nonradiative transition to level 2

where the first term on the right-hand side represents the stimulated transitions between levels 1 and 2, the second term represents spontaneous transition from level 3 to level 2, and the third term represents the spontaneous transitions from level 2 to level 1 The quantity W_l is proportional to the Einstein coefficient B_{21} and to the energy density associated with the lasing transition 2 → 1 [see Eq (3 2-17)] Further T_{21} represents the net spontaneous relaxation rate from level 2 to level 1 If this transition is predominantly radiative then $T_{21} \approx A_{21}$, where A_{21} represents the Einstein coefficient

Similarly,

$$\frac{dN_1}{dt} = W_p(N_3 - N_1) + W_l(N_2 - N_1) + N_2 T_{21} \qquad (4\ 2\text{-}5)$$

where the first two terms represent the stimulated transitions between levels 1 and 3 and levels 1 and 2, respectively, and the last term represents the spontaneous transitions from level 2 to level 1 One can immediately see that

$$\frac{dN_1}{dt} + \frac{dN_2}{dt} + \frac{dN_3}{dt} = 0 \qquad (4\ 2\text{-}6)$$

which is consistent with Eq (4 2-1)

Equations (4 2-2), (4 2-4), and (4 2-5) are referred to as the rate equations and give the rate of change of populations of the three levels in a three-level laser system in terms of W_p and W_l At steady state, we can set the time derivatives to zero and, using Eq (4 2-1), solve for N_1, N_2, and N_3 Thus from Eq (4 2-2) we obtain

$$N_3 = \frac{W_p}{W_p + T_{32}} N_1 \qquad (4\ 2\text{-}7)$$

Similarly from Eq (4 2-4), we obtain, putting $dN_2/dt = 0$,

$$N_2 = \left(W_l + \frac{T_{32} W_p}{W_p + T_{32}}\right) \frac{N_1}{W_l + T_{21}} \qquad (4\ 2\text{-}8)$$

Using Eqs (4 2-1), (4 2-7), and (4 2-8), we obtain the expression for the population difference between the levels 2 and 1 as

$$\frac{N_2 - N_1}{N} = \frac{W_p(T_{32} - T_{21}) - T_{32}T_{21}}{3W_pW_l + 2W_pT_{21} + 2T_{32}W_l + T_{32}W_p + T_{32}T_{21}} \qquad (4\ 2\text{-}9)$$

From Eq (4 2-9) it follows that in order to obtain population inversion between levels 2 and 1, i e , for $N_2 - N_1$ to be positive, a

necessary condition is that

$$T_{32} > T_{21} \tag{4.2-10}$$

Since the relaxation times of atoms in levels 3 and 2 are inversely proportional to the corresponding relaxation rates, according to Eq (4.2-10), the lifetime of level 3 must be at least smaller than the lifetime of level 2 for attainment of population inversion. In addition, in order to achieve population inversion, a minimum pump power is required. This is given from Eq. (4.2-9) by

$$W_{pt} = \frac{T_{32}T_{21}}{T_{32} - T_{21}} \tag{4.2-11}$$

For obtaining population inversion, W_p is required to be greater than W_{pt}.

Equation (4.2-9) gives the population inversion as a function of W_p, W_l, and various parameters of the atomic system. For low laser powers, i.e., when W_l is very small compared to T_{21}, then we may neglect the W_l terms in Eq. (4.2-9) and assuming $T_{32} \gg T_{21}$, we obtain for the population difference

$$\frac{N_2 - N_1}{N} = \frac{W_p - T_{21}}{W_p + T_{21}} \tag{4.2-12}$$

Under this approximation for population inversion, we must have $W_p > T_{21}$ and the population inversion is independent of the energy corresponding to the laser transition. Thus, from Eqs. (3.2-21) and (3.2-22) it follows that the intensity of the beam grows exponentially.

If, on the other hand, the power at the laser transition frequency is very high, then we can write from Eq. (4.2-9),

$$\frac{N_2 - N_1}{N} \simeq \frac{W_p T_{32} - T_{32}T_{21}}{W_l(3W_p + 2T_{32})} \tag{4.2-13}$$

Under this approximation, the population inversion is inversely proportional to W_l and hence to the beam intensity [see Eq. (3.2-20)]. Thus α_ω is inversely proportional to I_ω and hence dI_ω/dz in Eq. (3.2-21) would be independent of I_ω. For such a case, the intensity of the beam grows only linearly with distance. Equation (4.2-12) can be rewritten in the form

$$\frac{N - \Delta N}{2} W_p = \frac{N + \Delta N}{2} T_{21} \tag{4.2-14}$$

where $\Delta N = N_2 - N_1$. Now, for a three-level laser system, one can write

$$N = N_1 + N_2 + N_3 \simeq N_1 + N_2$$

where we have assumed that the atoms from level 3 drop down to level 2 so quickly that level 3 is essentially unpopulated. Thus we can write

$$\frac{N - \Delta N}{2} \simeq N_1 \quad \text{and} \quad \frac{N + \Delta N}{2} \simeq N_2 \tag{4.2-15}$$

Consequently, in Eq. (4.2-14), the left-hand side represents the number of atoms per unit volume (in level 1) lifted to level 2 per unit time and the right-hand side represents the number of atoms per unit volume (in level 2) decaying spontaneously to level 1.

As an example of the three-level laser system, we consider the ruby laser.† For population inversion (i.e., for $N_2 > N_1$) we must have [see Eq. (4.2-12)]

$$W_p > T_{21} \simeq A_{21} = \frac{1}{t_{sp}}$$

Using the value of t_{sp} given in Eq. (3.2-31), we obtain the following value for the threshold pumping rate for population inversion

$$W_{pt} = 330 \text{ sec}^{-1} \tag{4.2-16}$$

We will now calculate the minimum amount of pump power to be spent to maintain population inversion. The rate at which atoms decay from the upper laser level 2 is given by $N_2 T_{21}$. For each atom lifted to level 2 one has to supply at least an amount of energy given by $h\nu_p$, where ν_p represents the average pump frequency. Hence in order to maintain N_2 atoms in the level 2, the minimum power P to be spent would be $N_2 h\nu_p / t_{sp}$ per unit volume of the active material where we have assumed $T_{21} \simeq 1/t_{sp}$. Thus, we have

$$P = \frac{N_2 h\nu_p}{t_{sp}} \tag{4.2-17}$$

Now since $(N_2 - N_1)_t \ll N$ [see Eq. (3.2-34)], we get from Eq. (4.2-15)

$$N_2 \simeq \frac{N}{2} \tag{4.2-18}$$

Thus the threshold pumping power per unit volume required to maintain population inversion in a three-level laser system is

$$P_t = \frac{N}{2} \frac{h\nu_p}{t_{sp}} \tag{4.2-19}$$

† See Section 9.3.

Taking the average pump frequency as $\nu_p = 6.25 \times 10^{14}$ Hz (which is averaged over the green and violet absorption bands), we obtain,

$$P_t \approx \frac{6.7 \times 10^{-34} \times 6.25 \times 10^{14} \times 1.6 \times 10^{19}}{2 \times 3 \times 10^{-3}}$$

$$\approx 1100 \text{ W/cm}^3 \tag{4.2-20}$$

If we assume that the efficiency of the pumping source is 25% and also that only 25% is absorbed in passage through the ruby rod, then the electrical threshold power comes out to be about 18 kW/cm³ of the active material. This is consistent with the threshold powers determined experimentally.

Under pulsed operation, if we assume that the pumping pulse is much shorter than the lifetime of level 2, then the atoms excited to the upper laser level do not appreciably decay during the duration of the pulse and the threshold pump energy would be

$$U_{pt} = \frac{N}{2} h\nu_p \tag{4.2-21}$$

per unit volume of the active medium. For the case of ruby laser, with the above efficiencies of pumping and absorption, one obtains

$$U_{pt} \approx 54 \text{ J/cm}^3 \tag{4.2-22}$$

It may be noted here that even though the ruby laser is a three-level laser system, because of various other factors mentioned below it does operate with not too large a pumping power. Thus, for example, the absorption band of ruby crystal is very well matched to the emission spectrum of available pump lamps so that the pumping efficiency is quite high. Also, most of the atoms pumped to level 3 drop down to level 2, which has a very long (3×10^{-3} sec) lifetime which is nearly radiative. In addition the linewidth of the laser transition is also very narrow.

4.3 The Four-Level System

The four-level laser system is shown in Fig. 4.2. Level 1 is the ground level and levels 2, 3, and 4 are excited levels of the system. Atoms from level 1 are excited by a pump light to level 4, from which the atoms decay very rapidly through some nonradiative transition to level 3. Level 3 is a metastable level having a long lifetime ($\sim 10^{-3}$ sec). This level forms the upper laser level and level 2 forms the lower laser level. The lower laser level must have a very short lifetime so that the incoming atoms from

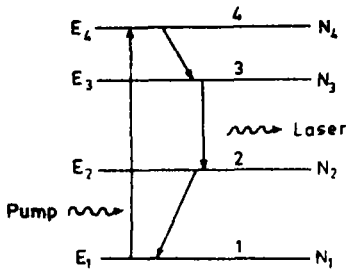



Fig 4 2 The energy levels in a four-level laser system Level 1 is the ground level and levels 2 3, and 4 are the excited levels The pump lifts atoms from level 1 to level 4 from where they make a fast transition to level 3 which has a slow relaxation rate Laser action is obtained between levels 3 and 2 when N_3 becomes greater than N_2 The atoms dropping to E_2 make another fast nonradiative transition to the ground level 1

level 3 relax down immediately from level 2 to level 1, ready for being pumped to level 4 If the rate of relaxation of atoms from level 2 to level 1 is faster than the rate of arrival of atoms into level 2, one can obtain population inversion between levels 3 and 2 even for very small pump powers The amount of population inversion between levels 3 and 2 would depend on the pumping rate

The level 4 could be a collection of a large number of levels or a broad level Then an optical pump emitting radiation over a band of frequencies can efficiently excite atoms from level 1 to level 4 Also, level 4 cannot be the upper laser level because the upper laser level is required to be narrow In addition the lower laser level 2 is required to be sufficiently above the ground level 1 so that at ordinary temperatures, the population of level 2 is negligible If the level 2 is not sufficiently above level 1 then the pumping power required for achieving population inversion would be much more At the same time the population of the level 2 may also be lowered by lowering the temperature of the system

Similar to the analysis carried out for the three-level system, we first write down the rate equations for the various levels of the four-level system Let N_1 N_2, N_3 and N_4 represent the population per unit volume of the levels 1, 2, 3, and 4, respectively The change in population of level 4 is described by the following equation

$$\frac{dN_4}{dt} = W_p(N_1 - N_4) - T_4 N_4 \qquad (4\ 3\text{-}1)$$

where as before, $W_p(N_1 - N_4)$ represents the net rate of stimulated transitions occurring between levels 1 and 4 caused by the pump beam, T_4 is the net relaxation rate from level 4 to the lower levels 3, 2, and 1, i e ,

$$T_4 = T_{43} + T_{42} + T_{41} \qquad (4\ 3\text{-}2)$$

Here the T s represent the total relaxation rates (i e , both radiative and nonradiative) Also T_{43} is usually much greater than T_{42} and T_{41}, so that

most of the atoms pumped to level 4 drop down to level 3 The first term in Eq (4 3-1) represents the rate at which atoms are being pumped from level 1 to level 4 and the second term represents the rate at which atoms decay from level 4

Similarly, the rate equation for N_3 would be

$$\frac{dN_3}{dt} = T_{43}N_4 + W_l(N_2 - N_3) - T_3N_3 \tag{4 3-3}$$

where $T_3 = T_{32} + T_{31}$ The first term represents the rate at which atoms drop down from level 4 into level 3, the second term represents the rate of stimulated transitions from level 2 to level 3 due to the presence of laser radiation, and the third term represents the rate of loss of atoms from level 3 to levels 2 and 1 through spontaneous transitions

Similarly, the rate equation for N_2 and N_1 would be

$$\frac{dN_2}{dt} = T_{42}N_4 + W_l(N_3 - N_2) - T_{21}N_2 + T_{32}N_3 \tag{4 3-4}$$

$$\frac{dN_1}{dt} = W_p(N_4 - N_1) + T_{41}N_4 + T_{31}N_3 + T_{21}N_2 \tag{4 3-5}$$

with similar physical interpretations as in Eq (4 3-3) It can be seen from Eqs (4 3-1)–(4 3-5) that

$$\frac{dN_1}{dt} + \frac{dN_2}{dt} + \frac{dN_3}{dt} + \frac{dN_4}{dt} = 0 \tag{4 3-6}$$

which is consistent with the requirement that the total number of atoms

$$N = N_1 + N_2 + N_3 + N_4 \tag{4 3-7}$$

must be a constant

At steady state we must have

$$\frac{dN_1}{dt} = \frac{dN_2}{dt} = \frac{dN_3}{dt} = \frac{dN_4}{dt} = 0 \tag{4 3-8}$$

Using this, we get from Eq (4 3-1)

$$N_4 = \frac{W_p}{2W_p + T_4}(N - N_2 - N_3) \tag{4 3-9}$$

where we have used Eq (4 3-7) Using Eqs (4 3-7), (4 3-9) and Eqs

(4 3-3), (4 3-4) with Eq (4 3-8), we obtain

$$N_3\left[(W_l + T_3)\left(\frac{2W_p + T_4}{W_p}\right) + T_{43}\right] + N_2\left[T_{43} - \frac{W_l(2W_p + T_4)}{W_p}\right] = T_{43}N \quad (4\ 3\text{-}10)$$

$$N_3\left[T_{32} - (T_{32} + W_l)\left(\frac{2W_p + T_4}{W_p}\right)\right] + N_2\left[(T_{21} + W_l)\left(\frac{2W_p + T_4}{W_p}\right) + T_{42}\right] = NT_{42} \quad (4\ 3\text{-}11)$$

We have now to solve the above two equations to obtain the population difference between levels 3 and 2, namely, $N_3 - N_2$ By simple algebraic manipulations we obtain

$$\frac{N_3 - N_2}{N} = (T_{21}T_{43} - T_3T_{42} - T_{32}T_{43}) \times \left\{T_{43}(T_{21} + T_{32}) + T_3T_{42} + \frac{(2W_p + T_4)}{W_p}T_3T_{21} + W_l\left[2(T_{42} + T_{43}) + \left(\frac{2W_p + T_4}{W_p}\right)(T_{31} + T_{21})\right]\right\}^{-1} \quad (4\ 3\text{-}12)$$

Equation (4 3-12) gives the steady state population difference as a function of the pump power the laser power, and the lifetimes of the various states involved in the four-level system

In most four-level systems, the atoms from level 4 relax primarily to level 3 and hence $T_{42}, T_{41} \ll T_{43}$ Also the atoms from level 3 relax primarily to level 2 i e $T_3 = T_{32}$ Under these approximations, we see that in order to obtain a population inversion, i e , in order to obtain a positive value of $N_3 - N_2$, we must have

$$T_{21} > T_{32} \quad (4\ 3\text{-}13)$$

i e , the rate at which atoms relax from level 2 to level 1 must be greater than the rate at which atoms relax from level 3 to level 2 Under such a condition, the creation of population inversion between levels 3 and 2 is independent of the pumping power but the magnitude of the population inversion depends on W_p

At and below threshold of oscillations, $W_l = 0$ and we can obtain an approximate equation for the population difference from Eq (4 3-12), as

$$\frac{N_3 - N_2}{N} = \frac{(1 - \beta)W_pT_{43}/T_4T_3}{1 + [(1 + \beta) + 2T_3/T_{43}]W_pT_{43}/T_4T_3} \quad (4\ 3\text{-}14)$$

where

$$\beta = \frac{T_{32}}{T_{21}} + \frac{T_3 T_{42}}{T_{21} T_{43}} \qquad (4\ 3\text{-}15)$$

For good laser action one must have $T_3 \ll T_{43}$ and $T_{21} \gg T_{32}$ so that $\beta \to 0$ Also $T_4 \simeq T_{43}$ Under such an approximation, Eq (4 3-14) reduces to

$$\frac{N_3 - N_2}{N} = \frac{W_p/T_3}{1 + W_p/T_3} \qquad (4\ 3\text{-}16)$$

Comparing Eqs (4 2-12) and (4 3-16) one can see that inversion can be much more easily obtained in a four-level scheme as compared to a three-level scheme In fact this has been found to be true experimentally

An important example of a four-level laser is the Nd YAG laser† The values of various parameters for a 1 at % doped crystal (which corresponds to a population density of Nd^{3+} ions of 6×10^{19} ions/cm^3) are

$$\Delta\nu = 1\ 95 \times 10^{11}\ \text{Hz}, \qquad \nu = 2\ 83 \times 10^{14}\ \text{Hz}$$
$$t_{sp} = 2\ 3 \times 10^{-4}\ \text{sec}, \qquad \lambda = \frac{1\ 06 \times 10^{-4}}{1\ 82} \simeq 0\ 58 \times 10^{-4}\ \text{cm} \qquad (4\ 3\text{-}17)$$

where we have used the fact that the refractive index of Nd:YAG is about 1 82 If we assume a cavity length of 7 cm and a loss per bounce of 5% then using Eq (3 2-32) we obtain

$$t_c \simeq 8 \times 10^{-9}\ \text{sec} \qquad (4\ 3\text{-}18)$$

Thus, we can use Eq (3 2-30) for calculating the threshold population inversion density

$$(N_3 - N_2)_t \simeq \frac{4 \times (3\ 14)^2}{(0\ 58)^3 \times 10^{-12}} \frac{1\ 95 \times 10^{11}}{2\ 83 \times 10^{14}} \frac{2\ 3 \times 10^{-4}}{8 \times 10^{-9}}$$
$$\simeq 4 \times 10^{15}\ \text{atom/cm}^3 \qquad (4\ 3\text{-}19)$$

Thus, the critical inversion density is much less than the total number of atoms per unit volume

Since $(N_3 - N_2)_t \ll N$, we obtain from Eq (4 3-16),

$$W_{pt} \simeq \frac{(N_3 - N_2)_t}{N} T_3 \qquad (4\ 3\text{-}20)$$

† In the Nd YAG laser the active medium consists of Nd^{3+} ions doped in yttrium aluminium garnet ($Y_3Al_5O_{12}$) For more details see Section 9 5

which for the present case is

$$W_{pt} = \frac{4 \times 10^{15}}{6 \times 10^{19}} \times \frac{1}{2.3 \times 10^{-4}} \simeq 0.3 \text{ sec}^{-1} \qquad (4.3\text{-}21)$$

which is much smaller than that obtained for the case of ruby laser [Eq (4.2-16)]

In order to calculate the critical power per unit volume required to maintain inversion, we again use Eq (4.2-17) with N_2 replaced by N_3 because level 3 is now the upper laser level Also, since in the four-level case, the lower laser level (namely, level 2) is almost empty (because of fast relaxation to level 1) $N_3 \simeq (N_3 - N_2)_t$ Thus if we assume that the pump has an average frequency of $\nu_p \simeq 4 \times 10^{14}$ Hz, then using the various parameters in Eq (4.2-17), we get

$$P_t \simeq \frac{4 \times 10^{15} \times 6.7 \times 10^{-34} \times 4 \times 10^{14}}{2.3 \times 10^{-4}}$$

$$\simeq 5 \text{ W/cm}^3 \qquad (4.3\text{-}22)$$

which is again much smaller than the case for ruby If we assume the same pump efficiency as in the case of ruby we obtain a threshold power of about 80 W/cm³

As another example of the four-level system, let us consider the helium–neon laser† The neon transition is inhomogeneously broadened and hence is Gaussian in shape Thus the threshold population inversion would be [see discussion after Eq (3.2-30)]

$$N_2 - N_1 = \frac{4\pi^2}{\lambda^3} \frac{\Delta\nu}{(\pi \ln 2)^{1/2}\nu} \frac{t_{sp}}{t_c} \qquad (4.3\text{-}23)$$

We consider the 1.15-μm transition in Ne for which we have the following values of the various parameters (Bennett 1962)

$$\lambda = 1.15\ \mu\text{m} \qquad \nu = 2.6 \times 10^{14} \text{ Hz}$$
$$\Delta\nu = 8 \times 10^8 \text{ Hz} \qquad t_{sp} = 10^{-7} \text{ sec} \qquad (4.3\text{-}24)$$

The cavity lifetime $t_c \simeq 1.6 \times 10^{-7}$ sec for a cavity length of 93.7 cm and a reflection loss of 2% Substituting these in Eq (4.3-23) we get for the threshold population inversion

$$N_2 - N_1 \simeq 3.4 \times 10^7 \text{ atom/cm}^3 \qquad (4.3\text{-}25)$$

In order to determine the power requirement since the excitation of Ne atoms is through resonant transfer of excitation from helium, ν_p in

† See Section 9.4 for details of the He–Ne laser

Eq (4 2-17) would be $\simeq 4\,8 \times 10^{15}$ Hz (corresponding to the 2^3S state of helium) Thus we obtain

$$P_t \simeq \frac{3\,4 \times 10^7 \times 6\,7 \times 10^{-34} \times 4\,8 \times 10^{15}}{10^{-7}}$$

$$\simeq 1\,1 \times 10^{-3}\ \text{W/cm}^3 \qquad (4\,3\text{-}26)$$

Thus if we have the laser gas in a volume of 100 cm³ (length of the discharge tube ~100 cm and area of cross section 1 cm²) the minimum power required to be spent is 110 mW

4 4 *Variation of Laser Power around Threshold*

In Sections 4 2 and 4 3, we had considered the three-level and four-level laser systems in detail and had obtained the pumping rate required to maintain population inversion in the laser transition In this section, we will consider a simple two-level system (see Fig 4 3) with the lower laser level having a very fast transition to other lower levels so that it is essentially unpopulated† We will write down the rate equations for the atomic population of the upper level and also for the number of photons in a cavity mode of the laser We would assume that only one mode has sufficient gain to oscillate and that the line is homogeneously broadened‡ We would obtain the number of photons in the cavity mode as a function of the pumping rate and study the variation of this photon number as the pumping rate passes through the threshold value We will also obtain the optimum output coupling in terms of the pump rate for maximum output power of the laser

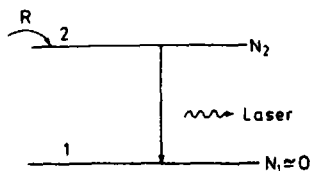



Fig 4 3 A two-level system The lower level is assumed to be unpopulated and the upper level has a population of N_2 The upper level is being pumped at a rate R

We assume that such a two-level system is placed inside an optical resonator§ and is being pumped by some means so that there is a net

† Such an analysis is valid for a four-level system with the lower laser level having an extremely small lifetime The general conclusions of this simple analysis remain valid even for a more complex three-level system for which a similar analysis can be made

‡ The analysis is valid for a homogeneously broadened line as we will be assuming that the same induced rate applies to all atoms

§ See Chapter 6 for a detailed discussion of optical resonators

pumping rate per unit volume denoted by R on the upper laser level, namely level 2 The atoms from the upper laser level decay both by spontaneous and stimulated transitions The energy in the cavity mode will be described by the number of photons present in the cavity mode

We show in Chapter 6 that an optical resonator having dimensions that are much greater than the wavelength of light have a very large number of closely spaced modes of oscillation Thus the number of modes having their frequencies inside the atomic linewidth is enormous ($\sim 10^8$) (see Section 6 2) The spontaneous emission which appears when the atom drops from level 2 to level 1 can appear in any one of the large number of cavity modes In fact, the total transition rate of the atom in all these different cavity modes is the same as the net spontaneous radiative relaxation rate on the $2 \rightarrow 1$ transition Also, as we will show in Section 8 5 this spontaneous emission rate is independent of the number of photons present in a cavity mode Thus, the total spontaneous emission rate from the upper level would be $N_2 A_{21}$, where A_{21} is the Einstein coefficient In addition to this radiative relaxation, one could also have some nonradiative relaxation Thus the net relaxation rate would be $N_2 T_{21}$ where as before $T_{21} = A_{21} + S_{21}$, S_{21} being the nonradiative relaxation rate

Out of the extremely large number of cavity modes lying under the linewidth of the atomic transition, only one or a few modes would, in general have significantly lower loss and hence a lower threshold Thus only the mode with highest gain would oscillate (for a homogeneous transition) as the pumping rate is increased beyond threshold

In order to evaluate the stimulated transition rate into a mode in which n photons are already present we refer to Section 3 3 where we obtained the stimulated transition rate in the presence of a monochromatic field of radiation energy density u_ω as [see Eq (3 3-45)]

$$F_{21} = W_{21} V = \frac{\pi^2 c^3}{\hbar \omega^3} A_{21} u_\omega g(\omega) N_2 V \qquad (4\ 4\text{-}1)$$

where as before N represents the number of atoms per unit volume in level 2 and V represents the volume of the resonator cavity If we assume that there are n photons in this mode then

$$u_\omega = \frac{n \hbar \omega}{V} \qquad (4\ 4\text{-}2)$$

where V is the volume of the cavity Also, if we assume that this mode lies at the center of the atomic line then (see Section 3 5)

$$g(\omega_0) = \frac{\tau_0}{\pi} = \frac{2}{\pi\, \Delta\omega} \qquad (4\ 4\text{-}3)$$

and one obtains

$$F_{21} = \frac{\lambda^3 A}{4\pi^2 V} \frac{\nu}{\Delta\nu} N_2 n V$$
$$= K N_2 n \qquad (4\ 4\text{-}4)$$

where

$$K = \frac{\lambda^3}{4\pi^2} A \frac{\nu}{\Delta\nu} \qquad (4\ 4\text{-}5)$$

Thus, Eq (4 4-4) gives the transition rate from level 2 to level 1 due to stimulated transitions Consequently, the net rate of change of the population of level 2 would be given by

$$\frac{dN_2 V}{dt} = -KnN_2 - N_2 V T_{21} + RV \qquad (4\ 4\text{-}6)$$

where the first term represents the number of stimulated transitions

We have now to write an equation for the rate of change of photon number n in the mode of the cavity The photon number n changes because of (a) the stimulated transitions from level 2, (b) the spontaneous transitions which occur from level 2 and which emit a photon in this particular mode of the cavity, and (c) the losses in the cavity like scattering, output coupling through the mirror, etc

The contribution due to stimulated transition is exactly the negative of the first term on the right-hand side of Eq (4 4-6), namely, KN_2n We show in Chapter 8 that the spontaneous emission rate into a particular mode is exactly the same as the stimulated emission rate into that mode when there is only one photon in that mode Hence, this spontaneous rate can be written as† KN_2 If we represent by t_c the cavity lifetime then the rate of loss of photons would be n/t_c

Hence, we can write for the net rate of change of the photon number in the cavity mode

$$\frac{dn}{dt} = K(n + 1)N_2 - \frac{n}{t_c} \qquad (4\ 4\text{-}7)$$

We have now to solve Eqs (4 4-6) and (4 4-7) Observe that Eqs (4 4-6) and (4 4-7) are nonlinear because of the product term n N_2

† The quantity KN_2 represents the number of spontaneous emissions (per unit time) in the entire volume of the cavity into a particular mode Notice that K is independent of the volume This at first sight may appear inconsistent however one must remember that the total number of modes (lying inside the linewidth of the cavity medium) is proportional to the volume [see Eq (8 5-30)] and hence the total number of spontaneous emissions inside the cavity will be proportional to volume

We first solve the above rate equations under steady state condition, when we can set the time derivatives to zero. Considering Eq. (4.4-7), and putting $dn/dt = 0$, we obtain

$$N_2 = \left(\frac{n}{n+1}\right)\frac{1}{Kt_c} \tag{4.4-8}$$

The above equation represents the steady state value† of N_2. Substituting this value of N_2 in Eq. (4.4-6) we obtain for the steady state (i.e., $dN_2/dt = 0$),

$$\frac{K}{VT_{21}}n^2 + n\left(1 - \frac{R}{R_t}\right) - \frac{R}{R_t} = 0 \tag{4.4-9}$$

where

$$R_t = T_{21}/Kt_c \tag{4.4-10}$$

The solution of Eq. (4.4-9), which gives positive values of n is

$$n = \frac{VT_{21}}{2K}\left\{\left(\frac{R}{R_t} - 1\right) + \left[\left(\frac{R}{R_t} - 1\right)^2 + \frac{4R}{R_t}\frac{K}{VT_{21}}\right]^{1/2}\right\} \tag{4.4-11}$$

The above equation gives the variation of the photon number n (in the cavity mode) as a function of the pumping rate R. If we assume that the relaxation from level 2 is purely radiative, then $T_{21} \simeq A_{21} = 1/t_{sp}$. For typical values of various parameters in, for example, Nd–YAG laser system (see Section 4.3), one obtains

$$\frac{K}{VT_{21}} \simeq 10^{-11} \tag{4.4-12}$$

For such small values of K/VT_{21}, if R/R_t is assumed to be slightly different from 1, we may make a binomial expansion and readily obtain the following approximate formulas

$$n \simeq \frac{R}{R_t}\left(1 - \frac{R}{R_t}\right)^{-1} \qquad \text{for } \frac{R}{R_t} < 1 - \Delta \tag{4.4-13a}$$

$$n \simeq \frac{VT_{21}}{K}\left(\frac{R}{R_t} - 1\right) \qquad \text{for } \frac{R}{R_t} > 1 + \Delta \tag{4.4-13b}$$

† For the emission to be coherent, n must be much greater than 1 and N_2 would have a value very close to $1/Kt_c$, which is the threshold value of N_2 calculated in Section 3.2 for laser action (in which we had neglected the contribution due to spontaneous emissions). Indeed, in Eq. (3.2-29) if we neglect N_1 and use Eqs. (4.4-3) and (4.4-5), we would get $N_2 \simeq (Kt_c)^{-1}$.

where Δ should be much greater than $(2K/VT_{21})^{1/2}$ Further,

$$n = \left(\frac{VT_{21}}{K}\right)^{1/2} \quad \text{for} \quad \frac{R}{R_t} = 1 \tag{4 4-14}$$

Figure 4 4 shows the variation of the photon number n as a function of normalized pumping rate, R/R_t, for the value of $K/T_{21}V$ given by Eq (4 4-12) Notice the extremely rapid increase in n as the pumping rate

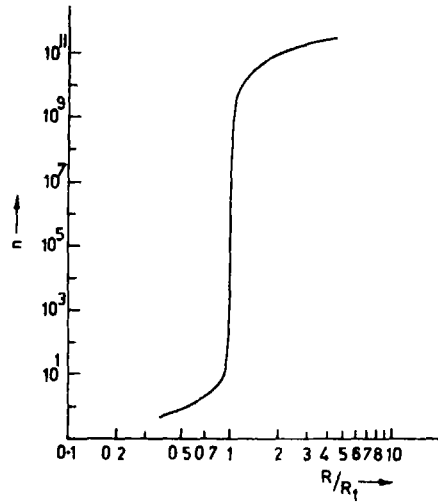



Fig 4 4 The variation of the cavity photon number n as a function of the normalized pumping rate R/R_t Observe that for $R/R_t < 1$ $n \sim 1$ and for $R/R_t > 1$ $n \sim 10^{11}$ $R = R_t$ corresponds to the threshold pumping rate

passes through the value R_t For example, for $R/R_t = 0\,9999$, $n = 9989$ and for $R/R_t = 1\,0001$, $n = 1\,001 \times 10^7$ In fact, one can show from Eq (4 4-11) that

$$\left.\frac{dn}{d(R/R_t)}\right|_{R=R_t} = \frac{VT_{21}}{2K} \simeq 5 \times 10^{10} \tag{4 4-15}$$

Thus for $n \gg 1$, R_t $(= T_{21}/Kt_c)$ represents the threshold pumping rate—consistent with the findings of Section 3 2

In an actual laser cavity, when the pumping rate is below threshold, the number of photons in any of the possible cavity modes is very small (~ 1) As one approaches the threshold, the number of photons in the preferred cavity mode (having higher gain and lower cavity losses) increases at a tremendous rate and, at and above threshold, the number of photons in the oscillating cavity mode (and hence the power output) is extremely large, the number of photons in other cavity modes which do not have a net (round trip) gain of unity remains orders of magnitude smaller

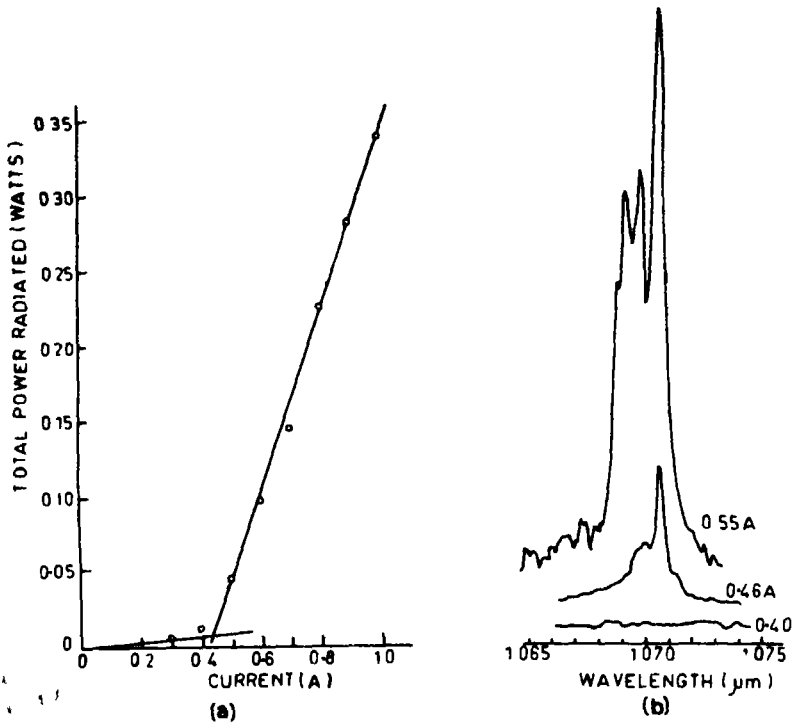



Fig 4 5 (a) The variation of the total power radiated in watts as a function of the curre a typical semiconductor laser The threshold current corresponds to 0 43 A (which in particular example corresponds to a current density of 1000 A/cm^2) Observe the steep in the output power as one passes through the threshold (b) The emission spectrum of laser output for different currents Observe the sharpening of the spectrum at a curren 0 46 A which implies the onset of laser oscillation (After Nuese *et al* 1976)

Apart from the sudden increase in the number of photons as (passes through threshold, the output of the laser changes rapidly from incoherent to a coherent nature The output becomes an almost p sinusoidal wave apart from small amplitude and phase fluctuations cau by the ever-present random spontaneous emissions† These spontane emissions determine the ultimate linewidth of the output laser beam Section 6 4, we derive an expression for this ultimate linewidth of a la oscillator

† In an actual laser system the ultimate purity of the output beam is restricted becaus the mechanical vibrations in the laser mirrors temperature fluctuations etc —see Section 6 4

Figure 4 5 shows a typical example of the variation of output power versus pumping in a semiconductor laser† as the pumping passes through the threshold region Notice the steep rise in output power near threshold

4.5 Optimum Output Coupling

In the earlier sections we calculated the steady state photon number in a cavity mode as a function of the pump rate In order to get a useful output laser beam, one usually uses a partially transparent mirror at one end of the resonator which couples out the laser beam In this section, we calculate the optimum value of the output coupling to obtain a maximum output power

The existence of an optimum value of output coupling can be understood as follows If the output coupling is made too large by increasing the transmittance of one of the mirrors, then the net losses in the cavity become too much and the laser may not oscillate at all On the other hand, if the transmittance of the mirror is made too small, then again the output power becomes too little Thus, for a given pumping rate there is an optimum value of the external coupling for maximum output power

Equation (4 4-11) gives the number of photons present in the cavity The rate at which the photons are lost from the mode is given by n/t_c [see Eq (4 4-7)] Out of this, a fraction is due to the internal losses and another part is due to the output coupling These we represent by n/t_i and n/t_e, respectively Thus

$$\frac{1}{t_c} = \frac{1}{t_i} + \frac{1}{t_e} \tag{4 5-1}$$

The output laser power would be given by

$$P_e = \frac{nh\nu}{t_e} = \frac{h\nu}{t_e}\frac{VT_{21}}{K}\frac{R - R_t}{R_t} \tag{4 5-2}$$

where we have used Eq (4 4-13b) for the number of photons in the cavity We now substitute the value of R_t from Eq (4 4-10) and use Eq

† See Section 9 8 and the Nobel lecture by Basov in Part III of the book

(4 5-1) to obtain

$$P_e = \frac{h\nu t_i V}{t_e + t_i}\left(R - \frac{T_{21}}{Kt_e} - \frac{T_{21}}{Kt_i}\right) \tag{4 5-3}$$

The maximum output power would correspond to $\partial P_e/\partial t_e = 0$, which gives from Eq (4 5-3),

$$\frac{1}{t_e} = \left(\frac{RK}{T_{21}t_i}\right)^{1/2} - \frac{1}{t_i} \tag{4 5-4}$$

Notice that the optimum output coupling is different for different pump powers Using this value of t_e in Eq (4 5-3), we get the maximum power output as

$$P_{max} = h\nu RV\left[1 - \left(\frac{T_{21}}{Kt_i R}\right)^{1/2}\right]^2 \tag{4 5-5}$$

If we use Eq (6 3-15) for the cavity lifetime, and assume that one of the mirrors is completely reflecting and the other has a transmittance T, then we obtain

$$\begin{aligned}\frac{1}{t_c} &= \frac{c\alpha_i}{n_0} - \frac{c}{2dn_0}\ln(1 - T)\\ &= \frac{1}{t_i} + \frac{1}{t_e}\end{aligned} \tag{4 5-6}$$

The above equation gives t_i and t_e in terms of other parameters of the cavity Also, for $T \ll 1$, we obtain

$$\frac{1}{t_e} \simeq \frac{cT}{2dn_0} \tag{4 5-7}$$

4 6 Laser Spiking

Throughout the present chapter, we have restricted ourselves to the steady state region of laser oscillation The rate equations that we have given above can also be used to study the behavior of the laser in the transient region As we had mentioned earlier, the rate equations for the atomic population and the photon number (or equivalently, the laser energy) are nonlinear, and a general analytical solution is not possible One has to resort to numerical solutions, we will describe here the main features of the behavior of the laser under transient conditions

When the pump is suddenly switched on to a value much above the threshold, the population inversion builds up and crosses the threshold

value The inversion can go beyond the threshold value under transient conditions because the photon number in the cavity begins to build and takes some time before reaching the threshold value, Since the inversion was much beyond threshold, the photon number would increase much beyond the steady state value and in the process deplete the upper level Since the photon number is higher than the steady state value, the inversion is brought below threshold, reducing the gain and the oscillation begins to die out The pump now builds up the inversion and again the whole cycle is repeated The net result is a series of bursts of laser action which is referred to as spiking For the parameters found in most solid state lasers, the spiking dies down only slowly with time Sometimes mechanical vibrations may also induce the spiking behavior Figure 9 11 shows the spiking in a typical ruby laser

5

Semiclassical Theory of the Laser

5.1. Introduction

The present chapter deals with the semiclassical theory of the laser as developed by Lamb (1964) In this analysis, we will treat the electromagnetic field classically with the help of Maxwell's equations and the atom will be treated using quantum mechanics We will consider a collection of two-level atoms placed inside an optical resonator The electromagnetic field of the cavity mode produces a macroscopic polarization of the medium This macroscopic polarization is calculated using quantum mechanics The polarization then acts as a source for the electromagnetic field in the cavity Since this field must be self-consistent with the field already assumed, one gets, using this condition the amplitude and frequencies of oscillation We will obtain explicit expressions for the real and imaginary parts of the electric susceptibility of the medium The real part is responsible for additional phase shifts due to the medium and leads to the phenomenon of mode pulling On the other hand, the imaginary part of the susceptibility is responsible for loss or gain due to the medium Under normal conditions, the population of the upper level is less than that of the lower level and the medium adds to the losses of the cavity In the presence of population inversion, the medium becomes an amplifying medium, however, a minimum population inversion is necessary to sustain oscillations in the cavity We will show that in the first-order theory, the electric field in the cavity can grow indefinitely, but using a third-order theory we would show that the field would indeed saturate rather than growing indefinitely

Since the analysis is semiclassical in nature, the effects of spontaneous emission do not appear Thus, the analysis does not give the ultimate linewidth of the laser oscillator which is caused by spontaneous emissions

5 2. Cavity Modes

We consider a laser cavity with plane mirrors at $z = 0$ and $z = L$ (see Fig 5 1) The electromagnetic radiation inside the cavity can be

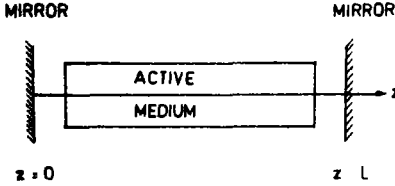



Fig 5 1 A plane parallel resonator bounded by a pair of plane mirrors facing each other The active medium is placed inside the resonator

described by Maxwell's equations, which in the MKS system of units are

$$\nabla \times \mathscr{E} = -\frac{\partial \mathscr{B}}{\partial t} \tag{5 2-1}$$

$$\nabla \times \mathscr{H} = \mathbf{J} + \frac{\partial \mathscr{D}}{\partial t} \tag{5 2-2}$$

$$\nabla \cdot \mathscr{D} = \rho \tag{5 2-3}$$

$$\nabla \cdot \mathscr{B} = 0 \tag{5 2-4}$$

where ρ represents the charge density and $\mathbf{J}$ the current density, $\mathscr{E}$, $\mathscr{D}$, $\mathscr{B}$ and $\mathscr{H}$ represent the electric field, electric displacement, magnetic induction, and magnetic field respectively Inside the cavity we may assume

$$\rho = 0 \tag{5 2-5}$$

$$\mathscr{B} = \mu_0 \mathscr{H} \tag{5 2-6}$$

$$\mathscr{D} = \epsilon_0 \mathscr{E} + \mathbf{P} \tag{5 2-7}$$

$$\mathbf{J} = \sigma \mathscr{E} \tag{5 2-8}$$

where $\mathbf{P}$ is the polarization, σ the conductivity, and ϵ_0 and μ_0 are the dielectric permittivity and magnetic permeability of free space It will be seen that the conductivity term leads to the medium being lossy which implies attenuation of the field, we will assume that other losses like those due to diffraction, finite transmission at the mirrors, etc are taken into account in σ Now,

$$\nabla \times (\nabla \times \mathscr{E}) = -\mu_0 \frac{\partial}{\partial t}(\nabla \times \mathscr{H}) = -\mu_0 \frac{\partial \mathbf{J}}{\partial t} - \mu_0 \frac{\partial^2 \mathscr{D}}{\partial t^2} \tag{5 2-9}$$

or

$$\nabla \times (\nabla \times \mathscr{E}) + \mu_0 \sigma \frac{\partial \mathscr{E}}{\partial t} + \epsilon_0 \mu_0 \frac{\partial^2 \mathscr{E}}{\partial t^2} = -\mu_0 \frac{\partial^2 \mathbf{P}}{\partial t^2} \tag{5 2-10}$$

If we assume the losses to be small and the medium to be dilute, we may neglect the second term on the left-hand side and the term on the right-hand side of the above equation to approximately obtain

$$\nabla \times (\nabla \times \mathscr{E}) + \epsilon_0 \mu_0 \frac{\partial^2 \mathscr{E}}{\partial t^2} = 0 \qquad (5.2\text{-}11)$$

Further since $\mathbf{P}$ is small, Eq (5 2-3) gives

$$0 = \nabla \cdot \mathscr{D} \approx \epsilon_0 \nabla \cdot \mathscr{E} \qquad (5.2\text{-}12)$$

Thus

$$\nabla \times \nabla \times \mathscr{E} = -\nabla^2 \mathscr{E} + \nabla(\nabla \cdot \mathscr{E}) \approx -\nabla^2 \mathscr{E} \qquad (5.2\text{-}13)$$

or

$$\nabla \times \nabla \times \mathscr{E} \approx -\frac{\partial^2 \mathscr{E}}{\partial z^2} \qquad (5.2\text{-}14)$$

where, in writing the last equation, we have neglected the x and y derivatives, this is justified when intensity variations in the directions transverse to the laser axis is small in distances $\sim \lambda$, which is indeed the case (see Chapter 6) Thus Eq (5 2-11) becomes

$$-\frac{\partial^2 \mathscr{E}}{\partial z^2} + \frac{1}{c^2} \frac{\partial^2 \mathscr{E}}{\partial t^2} = 0 \qquad (5.2\text{-}15)$$

where $c = (\epsilon_0 \mu_0)^{-1/2}$ represents the speed of light in free space If we further assume a specific polarization of the beam, Eq (5 2-15) becomes a scalar equation

$$\frac{\partial^2 \mathscr{E}}{\partial z^2} = \frac{1}{c^2} \frac{\partial^2 \mathscr{E}}{\partial t^2} \qquad (5.2\text{-}16)$$

which we solve by the method of separation of variables

$$\mathscr{E}(z, t) = Z(z) T(t) \qquad (5.2\text{-}17)$$

to obtain

$$\frac{1}{Z} \frac{d^2 Z}{dz^2} = \frac{1}{c^2} \frac{1}{T} \frac{d^2 T}{dt^2} = -K^2 \text{ (say)} \qquad (5.2\text{-}18)$$

Thus

$$Z(z) = A \sin(Kz + \theta) \qquad (5.2\text{-}19)$$

where the quantity K corresponds to the wave number At the cavity ends (i e , at $z = 0$ and $z = L$), the field [and hence $Z(z)$] will vanish,

giving

$$\theta = 0$$

and

$$K = \frac{n\pi}{L}, \qquad n = 1, 2, 3, \qquad (5\ 2\text{-}20)$$

We designate different values of K by K_n ($n = 1, 2, 3,$) The corresponding time dependence will be of the form

$$\cos \Omega_n t$$

where

$$\Omega_n = K_n c = \frac{n\pi c}{L} \qquad (5\ 2\text{-}21)$$

Thus the complete solution of Eq (5 2-16) would be given by

$$\mathscr{E}(z, t) = \sum_n A_n \cos(\Omega_n t) \sin(K_n z) \qquad (5\ 2\text{-}22)$$

If we next include the term describing the losses, we would have (instead of Eq 5 2-16)

$$\frac{\partial^2 \mathscr{E}}{\partial z^2} - \mu_0 \sigma \frac{\partial \mathscr{E}}{\partial t} = \frac{1}{c^2} \frac{\partial^2 \mathscr{E}}{\partial t^2} \qquad (5\ 2\text{-}23)$$

We assume the same spatial dependence ($\sim \sin K_n z$) and the time dependence to be of the form $\exp(i \Lambda_n t)$ to obtain

$$\Lambda_n^2 - \mu_0 \sigma c^2 i \Lambda_n - \Omega_n^2 = 0 \qquad (5\ 2\text{-}24)$$

or

$$\begin{aligned} \Lambda_n &= \tfrac{1}{2}[i \mu_0 \sigma c^2 \pm (-\mu_0^2 \sigma^2 c^4 + 4\Omega_n^2)^{1/2}] \\ &\simeq \pm \Omega_n + i\sigma/2\epsilon_0 \end{aligned} \qquad (5\ 2\text{-}25)$$

Thus the time dependence is of the form

$$\exp\left(-\frac{\sigma}{2\epsilon_0} t\right) e^{\pm i \Omega_n t} \qquad (5\ 2\text{-}26)$$

the first term describing the attenuation of the beam In the expression derived above, the attenuation coefficient does not depend on the mode number n, however, in general, there is a dependence on the mode number which we explicitly indicate by writing the time-dependent factor

as†

$$\exp\left(-\frac{\Omega_n}{2Q_n}t\right)e^{\pm i\Omega_n t} \tag{5 2-27}$$

where

$$Q_n = \frac{\epsilon_0}{\sigma}\Omega_n \tag{5 2-28}$$

represents the quality factor (see Section 6 3) Thus the solution of Eq (5 2-23) would be

$$\mathscr{E}(z, t) = \sum_n A_n \exp\left(-\frac{\Omega_n}{2Q_n}t\right)\cos(\Omega_n t)\sin(K_n z) \tag{5 2-29}$$

Finally, we try to solve the equation which includes the term involving the polarization

$$\frac{\partial^2\mathscr{E}}{\partial z^2} - \mu_0\sigma\frac{\partial\mathscr{E}}{\partial t} - \frac{1}{c^2}\frac{\partial^2\mathscr{E}}{\partial t^2} = \mu_0\frac{\partial^2 P}{\partial t^2} \tag{5 2-30}$$

[cf Eqs (5 2-16) and (5 2-23)] We assume $\mathscr{E}$ to be given by

$$\mathscr{E} = \tfrac{1}{2}\sum_n\{E_n(t)\exp[-i(\omega_n t + \phi_n(t))] + \text{c c}\}\sin K_n z \tag{5 2-31}$$

where c c stands for the complex conjugate (so that $\mathscr{E}$ is necessarily real), $E_n(t)$ and $\phi_n(t)$ are real slowly varying amplitude and phase coefficients, ω_n is the frequency of oscillation of the mode which may, in general, be slightly different from Ω_n We assume P to be of the form

$$P = \tfrac{1}{2}\sum_n\{P_n(t, z)\exp[-i(\omega_n t + \phi_n(t))] + \text{c c}\} \tag{5 2-32}$$

where $P_n(t, z)$ may be complex but is a slowly varying component of the polarization On substitution of $\mathscr{E}$ and P in Eq (5 2-30), we get‡ (after multiplying by c^2)

$$\Omega_n^2 E_n - i\left(\frac{\sigma}{\epsilon_0}\right)\omega_n E_n - 2i\omega_n \dot{E}_n - (\omega_n + \dot{\phi}_n)^2 E_n = \frac{\omega_n^2}{\epsilon_0}p_n(t) \tag{5 2-33}$$

$$p_n(t) = \frac{2}{L}\int_0^L P_n(t, z)\sin K_n z\, dz$$

† Because of the losses, the field in the cavity decays with time as $\exp(-\Omega_n t/2Q_n)$ and hence the energy decays as $\exp(-\Omega_n t/Q_n)$ Thus, the energy decays to $1/e$ of the value at $t = 0$ in a time $t_c = Q_n/\Omega_n$ which is referred to as the cavity lifetime (see also Section 6 3)

‡ Actually we have equated each Fourier component, this follows immediately by multiplying Eq (5 2-30) by $\sin K_m z$ and integrating from 0 to L.

where we have neglected small terms involving E_n, ϕ_n, $\ddot{P}_n$, $E_n\phi_n$, σE_n, $\sigma\phi_n$, $\phi_n p_n$ and p_n which are all of second order Now, since ω_n will be very close to Ω_n, we may write

$$\Omega_n^2 - (\omega_n + \phi_n)^2 \simeq 2\omega_n(\Omega_n - \omega_n - \phi_n) \tag{5 2-34}$$

Thus, equating real and imaginary parts of both sides of Eq (5 2-33), we get

$$(\omega_n + \phi_n - \Omega_n)E_n(t) = -\frac{1}{2}\frac{\omega_n}{\epsilon_0}\operatorname{Re}(p_n(t)) \tag{5 2-35}$$

$$E_n(t) + \frac{1}{2}\frac{\omega_n}{Q'_n}E_n(t) = -\frac{\omega_n}{2\epsilon_0}\operatorname{Im}(p_n(t)) \tag{5 2-36}$$

where

$$Q'_n = \frac{\epsilon_0\omega_n}{\sigma} \tag{5 2-37}$$

When $p_n = 0$, $\omega_n = \Omega_n$ and $E_n(t)$ will decrease exponentially with time—consistent with our earlier findings In general, if we define the susceptibility χ through the equation

$$p_n(t) = \epsilon_0\chi_n E_n(t) = \epsilon_0(\chi'_n + i\chi''_n)E_n(t) \tag{5 2-38}$$

where χ'_n and χ''_n represent, respectively, the real and imaginary parts of χ_n, then

$$\omega_n + \phi_n = \Omega_n - \tfrac{1}{2}\omega_n\chi'_n \tag{5 2-39}$$

and

$$E_n = -\frac{1}{2}\frac{\omega_n}{Q'_n}E_n - \tfrac{1}{2}\omega_n\chi''_n E_n(t) \tag{5 2-40}$$

The first term on the right-hand side represents cavity losses and the second term represents the effect of the medium filling the cavity It can be easily seen that if χ''_n is positive, then the cavity medium adds to the losses On the other hand if χ''_n is negative, the second term leads to gain † If

$$-\chi''_n = \frac{1}{Q'_n} \tag{5 2-41}$$

the losses are just compensated by the gain and Eq (5 2-41) is referred to as the threshold condition If $-\chi''_n > 1/Q'_n$, there would be a buildup of oscillation

† We will show in Section 5 3 that χ''_n is negative for a medium with a population inversion

From Eq (5.2-39), one may note that if we neglect the term ϕ_n, the oscillation frequency differs from the passive cavity frequency by $-\frac{1}{2}\omega_n\chi'_n$, which is known as the pulling term In order to physically understand the gain and pulling effects due to the cavity medium, we consider a plane wave propagating through the cavity medium If χ_n represents the electric susceptibility of the medium for the wave, then the permittivity ϵ of the medium would be

$$\epsilon = \epsilon_0 + \epsilon_0\chi_n = \epsilon_0(1 + \chi_n) \tag{5 2-42}$$

This implies that the complex refractive index of the cavity medium is

$$\left(\frac{\epsilon}{\epsilon_0}\right)^{1/2} = (1 + \chi_n)^{1/2} \simeq (1 + \tfrac{1}{2}\chi_n) = 1 + \tfrac{1}{2}\chi'_n + \frac{i}{2}\chi''_n \tag{5 2-43}$$

The propagation constant of the plane wave in such a medium would be

$$\begin{aligned}\beta &= \frac{\omega}{c}\left(\frac{\epsilon}{\epsilon_0}\right)^{1/2} \simeq \frac{\omega}{c}(1 + \tfrac{1}{2}\chi'_n) + \tfrac{1}{2}i\frac{\omega}{c}\chi''_n \\ &= \alpha + i\delta\end{aligned} \tag{5 2-44}$$

where

$$\alpha = \frac{\omega}{c}(1 + \tfrac{1}{2}\chi'_n), \qquad \delta = \frac{1}{2}\frac{\omega}{c}\chi''_n \tag{5 2-45}$$

Thus, a plane wave propagating along the z direction would have a z dependence of the form

$$e^{i\beta z} = e^{i\alpha z}e^{-\delta z} \tag{5 2-46}$$

In the absence of the component due to the laser transition $\chi'_n = \chi''_n = 0$ and the plane wave propagating through the medium undergoes a phase shift per unit length of ω/c The presence of the laser transition contributes both to the phase change and also to the loss or amplification of the beam Thus if χ''_n is positive, then δ is positive and the beam is attenuated as it propagates along the z direction On the other hand if χ''_n is negative, then the beam is amplified as it propagates through the medium As the response of the medium is stimulated by the field, the applied field and the stimulated response are phase coherent

In addition to the losses or amplification caused by the cavity medium, there is also a phase shift caused by the real part of the susceptibility χ'_n We will show in the next section that χ'_n is zero exactly at resonance, i e , if the frequency of the oscillating mode is at the center of the atomic line and it has opposite signs on either side of the line center This additional phase shift causes the frequencies of oscillation of

the optical cavity filled with the laser medium to be different from the frequencies of oscillation of the passive cavity (i e , the cavity in the absence of the laser medium) The actual oscillation frequencies are slightly pulled towards the center of the atomic line and hence the phenomenon is referred to as mode pulling

5 3 Polarization of the Cavity Medium

In the last section, we obtained equations describing the cavity field and the oscillation frequency of the cavity in terms of the polarization associated with the cavity medium In the present section, we consider a collection of two-level atoms and obtain an explicit expression for the macroscopic polarization (and hence the electric susceptibility) of the cavity medium in terms of the atomic populations in the two levels of the system The time-dependent Schrodinger equation is given by [see Eq (2 3-16)]

$$H\Psi = i\hbar\frac{\partial \Psi}{\partial t} \tag{5 3-1}$$

Where H is the quantum mechanical Hamiltonian of the system and Ψ represents the time-dependent wave function of the atomic system

Let H_0 represent the Hamiltonian of the atom and let $\psi_1(\mathbf{r})e^{-i\omega_1 t}$ and $\psi_2(\mathbf{r})e^{-i\omega_2 t}$ be the normalized wave functions associated with the lower level 1 and the upper level 2 respectively, of the atom Then

$$H_0\psi_1(\mathbf{r}) = E_1\psi_1(\mathbf{r}) \tag{5 3-2a}$$

$$H_0\psi_2(\mathbf{r}) = E_2\psi_2(\mathbf{r}) \tag{5 3-2b}$$

where $E_1 = \hbar\omega_1$ and $E_2 = \hbar\omega_2$ are the energies of the lower and the upper levels respectively The interaction of the atom with the electromagnetic field is described by

$$H' = -e\mathscr{E}.\mathbf{r} \tag{5 3-3}$$

which is assumed to be a perturbation on the Hamiltonian H_0, here $\mathscr{E}$ represents the electric field associated with the radiation In the presence of such an interaction we write the wave function as

$$\Psi(\mathbf{r}, t) = C_1(t)\psi_1(\mathbf{r}) + C_2(t)\psi_2(\mathbf{r}) \tag{5 3-4}$$

where $C_1(t)$ and $C_2(t)$ are time-dependent factors The physical significance of $C_1(t)$ and $C_2(t)$ is that $|C_1(t)|^2$ and $|C_2(t)|^2$ represent, respectively, the probability of finding the atom in the lower state ψ_1 and in the upper

state ψ_2 at time t Also, since we are considering a collection of N_v atoms per unit volume and each atom has a probability $|C_1(t)|^2$ of being found in the level ψ_1 at time t, the mean number of atoms per unit volume in the lower level 1, namely, N_1 would be

$$N_1 = N_v |C_1(t)|^2 \tag{5 3-5a}$$

Similarly, the mean number of atoms per unit volume in the level 2, N_2, would be

$$N_2 = N_v |C_2(t)|^2 \tag{5 3-5b}$$

Using Eqs (5 3-2) and (5 3-4), Eq (5 3-1) gives

$$i\hbar \sum_{n=1\,2} \dot{C}_n \psi_n = \sum_{n=1\,2} (\hbar\omega_n + H')C_n\psi_n \tag{5 3-6}$$

Multiplying both sides by ψ_1^* and integrating over spatial coordinates, one obtains

$$i\hbar \dot{C}_1 = E_1 C_1 + H'_{11} C_1 + H'_{12} C_2 \tag{5 3-7}$$

where

$$H'_{mn} = \int \psi_m^* H' \psi_n \, d\tau \tag{5 3-8}$$

Similarly by multiplying Eq (5 3-6) by ψ_2^* and integrating one obtains

$$i\hbar \dot{C}_2 = E_2 C_2 + H'_{22} C_2 + H'_{21} C_1 \tag{5 3-9}$$

But

$$H'_{11} = -e\mathscr{E} \int \psi_1^*(\mathbf{r})\mathbf{r}\psi_1(\mathbf{r}) \, d\tau = 0 \tag{5 3-10}$$

because $\mathbf{r}$ is an odd function Similarly

$$H'_{22} = 0 \quad \text{and} \quad H'_{12} = H'^*_{21} \tag{5 3-11}$$

Thus

$$\dot{C}_1(t) = \frac{1}{i\hbar}[E_1 C_1(t) + H'_{12} C_2(t)] \tag{5 3-12a}$$

and

$$\dot{C}_2(t) = \frac{1}{i\hbar}[E_2 C_2(t) + H'_{21} C_1(t)] \tag{5 3-12b}$$

In deriving the above equations, we have not considered any damping mechanism We wish to do so now by introducing phenomenological

damping factors Even though we are considering only two levels of the atomic system, the phenomenological damping factors take account of not only spontaneous transitions from the two levels but also, for example, collisions, etc This we do by rewriting Eqs (5 3-12a) and (5 3-12b) as

$$C_1(t) = -\frac{i}{\hbar} E_1 C_1 - \tfrac{1}{2}\gamma_1 C_1(t) - \frac{i}{\hbar} H'_{12} C_2(t) \qquad (5\ 3\text{-}13a)$$

$$C_2(t) = -\frac{i}{\hbar} E_2 C_2 - \tfrac{1}{2}\gamma_2 C_2(t) - \frac{i}{\hbar} H'_{21} C_1(t) \qquad (5\ 3\text{-}13b)$$

where γ_1 and γ_2 represent damping factors for levels 1 and 2, respectively In order to see more physically, we find that in the absence of any interaction when $H'_{12} = 0 = H'_{21}$ the solutions of Eqs (5 3-13a) and (5 3-13b) would be

$$C_1(t) = \text{const} \times \exp\left(-i\frac{E_1}{\hbar}t\right)e^{-(\gamma_1/2)t} \qquad (5\ 3\text{-}14a)$$

$$C_2(t) = \text{const} \times \exp\left(-i\frac{E_2}{\hbar}t\right)e^{-(\gamma_2/2)t} \qquad (5\ 3\text{-}14b)$$

Thus the probability of finding the atom in levels 1 and 2 (which are, respectively, proportional to $|C_1|^2$ and $|C_2|^2$) decays as $e^{-\gamma_1 t}$ and $e^{-\gamma_2 t}$, respectively Thus, the lifetimes of levels 1 and 2 are $1/\gamma_1$ and $1/\gamma_2$, respectively

We now define the following quantities†

$$\rho_{11} = C_1^* C_1, \qquad \rho_{12} = C_1 C_2^* \qquad (5\ 3\text{-}15)$$

$$\rho_{21} = C_1^* C_2, \qquad \rho_{22} = C_2^* C_2 \qquad (5\ 3\text{-}16)$$

Notice that ρ_{11} and ρ_{22} are nothing but the probabilities of finding the system in states 1 and 2, respectively Since we know the time dependence of C_1 and C_2 we can write down the time variation of the quantities ρ_{11} etc Thus,

$$\begin{aligned} \rho_{11} &= C_1 C_1^* + C_1 \dot{C}_1^* \\ &= -\left(i\frac{E_1}{\hbar} + \frac{\gamma_1}{2}\right)C_1 C_1^* - \frac{i}{\hbar} H'_{12} C_2 C_1^* \\ &\quad -\left(-i\frac{E_1}{\hbar} + \frac{\gamma_1}{2}\right)C_1^* C_1 + \frac{i}{\hbar} H'^*_{12} C_2^* C_1 \end{aligned}$$

† The four quantities ρ_{11} ρ_{12} ρ_{21} and ρ_{22} form the elements of what is known as the density matrix ρ, see Section 7 8

or

$$\rho_{11} = -\gamma_1\rho_{11} + \left(\frac{i}{\hbar} H'_{21}\rho_{12} + \text{c c}\right) \tag{5 3-17}$$

where c c represents the complex conjugate Similarly

$$\rho_{22} = -\gamma_2\rho_{22} - \left(\frac{i}{\hbar} H'_{21}\rho_{12} + \text{c c}\right) \tag{5 3-18}$$

$$\rho_{21} = -(i\omega_{21} + \gamma_{12})\rho_{21} + \frac{i}{\hbar} H'_{21}(\rho_{22} - \rho_{11}) \tag{5 3-19}$$

where

$$\omega_{21} = \frac{E_2 - E_1}{\hbar} \quad \text{and} \quad \gamma_{12} = \frac{\gamma_1 + \gamma_2}{2} \tag{5 3-20}$$

Now,

$$H'_{21} = -e\mathscr{E} \int \psi_2^* \mathbf{r} \psi_1 \, d\tau \tag{5 3-21}$$

and if we consider single mode operation with $\mathscr{E}$ along the x axis then

$$H'_{21} = -e\mathscr{E}_n \int \psi_2^* x \psi_1 \, d\tau = -\mathscr{E}_n \mathscr{P} \tag{5 3-22}$$

where

$$\mathscr{P} = e\int \psi_2^* x \psi_1 \, d\tau \tag{5 3-23}$$

and

$$\begin{aligned} \mathscr{E}_n &= \tfrac{1}{2}E_n(t)\exp\{-i[\omega_n t + \phi_n(t)]\}\sin K_n z + \text{c c} \\ &= E_n(t)\cos[\omega_n t + \phi_n(t)]\sin K_n z \end{aligned} \tag{5 3-24}$$

Further when the system is in the state $\Psi(t)$, the average dipole moment is given by

$$\begin{aligned} P_a &= e\int \Psi^* x \Psi \, d\tau \\ &= e\int (C_1^*\psi_1^* + C_2^*\psi_2^*) x (C_1\psi_1 + C_2\psi_2) \, d\tau \\ &= \mathscr{P}(\rho_{21} + \rho_{12}) \end{aligned} \tag{5 3-25}$$

where we have used the relation

$$\int \psi_1^* x \psi_2 \, d\tau = \int \psi_2^* x \psi_1 \, d\tau \tag{5 3-26}$$

which can always be made to satisfy by appropriate choice of phase factors Thus, in order to calculate P_a we must know ρ_{12} and its complex conjugate ρ_{21} We first present the first-order theory which will be followed by the more rigorous third-order theory

5 3.1 First-Order Theory

In the first-order theory, we assume $(\rho_{22} - \rho_{11})$ to be dependent only on z and to be independent of time†

$$\rho_{22} - \rho_{11} = N(z) \tag{5 3-27}$$

Referring to Eqs (5 3-5a) and (5 3-5b), we see that $N_v(\rho_{22} - \rho_{11})$ represents the difference (per unit volume) in the population of the upper and lower states Thus, Eq (5 3-19) becomes

$$\dot{\rho}_{21} = -(i\omega_{21} + \gamma_{12})\rho_{21} - \frac{i}{2\hbar}\mathscr{P}E_n(t)(e^{-i(\omega_n t + \phi_n)} + \text{c c}) \sin K_n z \, N(z) \tag{5 3-28}$$

where we have used Eqs (5 3-22) and (5 3-24) Now, if we neglect the second term on the right-hand side of the above equation, the solution would be of the form

$$\rho_{21}(t) = \rho_{21}^{(0)} \exp[-(i\omega_{21} + \gamma_{12})t] \tag{5 3-29}$$

We next assume the solution of Eq (5 3-28) to be of the above form with $\rho_{21}^{(0)}$ now depending on time On substitution in Eq (5 3-28), we obtain

$$\dot{\rho}_{21}^{(0)} = -i\frac{\mathscr{P}}{2\hbar} E_n(t) N(z) \sin K_n z \{\exp[-i(\omega_n - \omega_{21} + i\gamma_{12})t - i\phi_n] + \exp[i(\omega_n + \omega_{21} - i\gamma_{12})t - i\phi_n]\} \tag{5 3-30}$$

We neglect the time dependence of E_n and ϕ_n and integrate the above equation to get

$$\rho_{21}^{(0)} \simeq \frac{\mathscr{P}}{2\hbar} N(z) E_n(t) \sin K_n z \left\{ \frac{\exp[-i(\omega_n - \omega_{21} + i\gamma_{12})t - i\phi_n]}{\omega_n - \omega_{21} + i\gamma_{12}} - \frac{\exp[i(\omega_n + \omega_{21} - i\gamma_{12})t - i\phi_n]}{\omega_n + \omega_{21} + i\gamma_{12}} \right\} \tag{5 3-31}$$

† This will be justified in Section 5 3 2

We neglect the second exp term in the curly brackets in the above equation as it has very rapid variations and we obtain

$$\rho_{21}^{(0)} \simeq \frac{\mathscr{P}}{2\hbar} N(z) E_n(t) \sin K_n z \, e^{\gamma_{12} t} \frac{\exp\{-i[(\omega_n - \omega_{21})t + \phi_n(t)]\}}{(\omega_n - \omega_{21}) + i\gamma_{12}} \quad (5.3\text{-}32)$$

Thus,†

$$\rho_{21} \simeq \frac{\mathscr{P}}{2\hbar} \frac{N(z) \sin K_n z \, E_n(t)}{(\omega_n - \omega_{21}) + i\gamma_{12}} \exp[-i(\omega_n t + \phi_n)]$$

or

$$\rho_{21} \simeq \frac{\mathscr{P}}{2\hbar} \frac{N(z) \sin K_n z \, E_n(t)}{\Gamma_n} e^{-i\theta_n} \exp[-i(\omega_n t + \phi_n)] \quad (5.3\text{-}33)$$

where

$$\cos \theta_n = \frac{\omega_n - \omega_{21}}{\Gamma_n}, \qquad \sin \theta_n = \frac{\gamma_{12}}{\Gamma_n} \quad (5.3\text{-}34)$$

$$\Gamma_n \equiv [(\omega_n - \omega_{21})^2 + \gamma_{12}^2]^{1/2} \quad (5.3\text{-}35)$$

Thus, from Eq (5.3-25) we get for the average dipole moment per atom

$$P_a = \mathscr{P}(\rho_{21} + \rho_{12}) = \mathscr{P}(\rho_{21} + \text{c c})$$

If we assume that there are N_v atoms per unit volume in the cavity, then the macroscopic polarization would be given by

$$\begin{aligned} P &= N_v P_a \\ &= \frac{N_v \mathscr{P}^2}{2\hbar} \left[\frac{N(z) \sin K_n z}{\Gamma_n} E_n(t) e^{-i\theta_n} e^{-i(\omega_n t + \phi_n)} + \text{c c} \right] \end{aligned} \quad (5.3\text{-}36)$$

where we have used Eq (5.3-33) Comparing with Eq (5.2-32), we get

$$P_n(t, z) = \frac{N_v \mathscr{P}^2 E_n(t) N(z) \sin K_n z}{\hbar \Gamma_n} e^{-i\theta} \quad (5.3\text{-}37)$$

and

$$\begin{aligned} p_n(t) &= \frac{2}{L} \int_0^L P_n(t, z) \sin K_n z \, dz \\ &= \frac{\mathscr{P}^2 E_n(t)}{\hbar \Gamma_n} e^{-i\theta_n} \bar{N} N_v \end{aligned} \quad (5.3\text{-}38)$$

† The constant of integration in Eq (5.3-31) would have led to an exponentially decaying term in Eq (5.3-33)

where

$$\bar{N} = \frac{2}{L}\int_0^L N(z)\sin^2 K_n z\,dz \simeq \frac{1}{L}\int_0^L N(z)\,dz \qquad (5.3\text{-}39)$$

In writing the last step, we have assumed that $N(z)$ (which represents the population inversion density) varies slowly in an optical wavelength Comparing Eq (5 3-38) with Eq (5 2-38), we get

$$\chi'_n = \frac{\wp^2 \bar{N} N_v}{\hbar\epsilon_0}\frac{\omega_n - \omega_{21}}{(\omega_n - \omega_{21})^2 + \gamma_{12}^2} \qquad (5.3\text{-}40)$$

and

$$\chi''_n = -\frac{\wp^2 \bar{N} N_v}{\epsilon_0\hbar}\frac{\gamma_{12}}{(\omega_n - \omega_{21})^2 + \gamma_{12}^2} \qquad (5.3\text{-}41)$$

The above two equations represent the variation of the real and imaginary parts of the susceptibility with the mode frequency ω_n

In Fig 5 2 we have plotted the variation of χ'_n and χ''_n with ω_n for a medium with a population inversion (i e , $\bar{N} > 0$)

Substituting for χ''_n in Eq (5 2-40), we get

$$E_n(t) = \left[-\frac{\omega_n}{2Q'_n} + \frac{\omega_n \wp^2 \bar{N} N_v}{2\epsilon_0\hbar}\frac{\gamma_{12}}{(\omega_n - \omega_{21})^2 + \gamma_{12}^2}\right]E_n(t) \qquad (5.3\text{-}42)$$

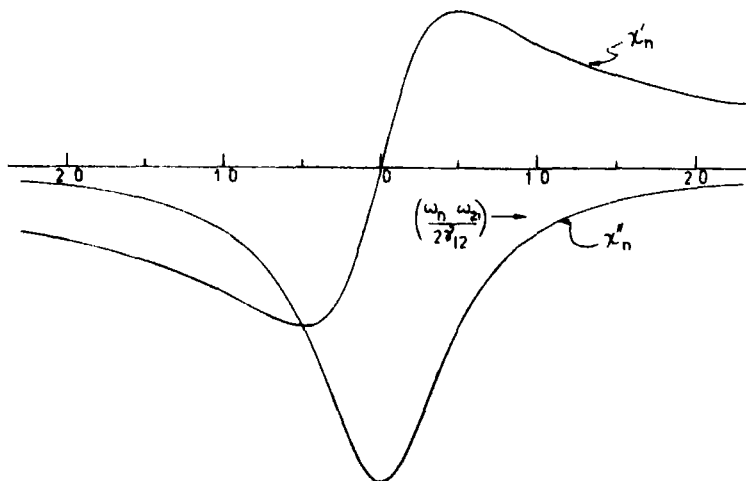



Fig 5 2 Variation of χ_n and χ_n which represent, respectively the real and imaginary parts of the electric susceptibility of the medium as a function of $\omega_n - \omega_{21}$ χ''_n is peaked at $\omega_n = \omega_{21}$ and thus maximum gain appears at $\omega_n = \omega_{21}$

Thus, for the amplitude to grow with time, the quantity inside the square brackets should be positive or

$$\frac{\mathcal{P}^2 \bar{N} N_v}{\epsilon_0 \hbar} \frac{\gamma_{12}}{(\omega_n - \omega_{21})^2 + \gamma_{12}^2} > \frac{1}{Q_n'} \quad (5\ 3\text{-}43)$$

When the two sides of the above inequality are equal, then the losses are exactly compensated by the gain due to the cavity medium and this corresponds to the threshold condition

The sign of the second term in the square brackets in Eq (5 3-42) depends on the sign of $\bar{N}$ It may be recalled that $\bar{N}$ is proportional to the population difference between the upper and the lower states† Thus if $\bar{N}$ is negative, i e , if there are more atoms in the lower level than in the upper level, then the second term contributes an additional loss On the other hand, if there is a population inversion between the levels 1 and 2 then $\bar{N}$ is positive and the medium acts as an amplifying medium In order that the mode may oscillate, the losses have to be compensated by the gain and this leads to the threshold condition, for which we must have

$$\bar{N}_t N_v = (N_2 - N_1)_t = \frac{\epsilon_0 \hbar}{\mathcal{P}^2 \gamma_{12}} \frac{(\omega_n - \omega_{21})^2 + \gamma_{12}^2}{Q_n'}$$

$$= \frac{\epsilon_0 \hbar}{\pi g(\omega) \mathcal{P}^2 Q_n'} \quad (5\ 3\text{-}44)$$

where the subscript t implies the threshold value and $g(\omega)$ represents the normalized line-shape function

$$g(\omega) = \frac{\gamma_{12}}{\pi} \frac{1}{(\omega_n - \omega_{21})^2 + \gamma_{12}^2} \quad (5\ 3\text{-}45)$$

which is identical to Eq (3 5-19) with $2t_{sp} = 1/\gamma_{12}$ Further [see Eqs (5 3-23), (3 4-4)]

$$\mathcal{P}^2 = e^2 \left[\int \psi_2^* x \psi_1 \, d\tau \right]^2 = \frac{e^2}{3} \left| \int \psi_2^* \mathbf{r} \psi_1 \, d\tau \right|^2$$

$$= \frac{\pi \epsilon_0 \hbar c^3}{\omega^3} A = \frac{\pi \epsilon_0 \hbar c^3}{\omega^3} \frac{1}{t_{sp}} \quad (5\ 3\text{-}46)$$

where t_{sp} is the spontaneous relaxation time of level 2 Substituting for $\mathcal{P}^2$ in Eq (5 3-46), we get an expression for $(N_2 - N_1)_t$ identical to Eq (3 2-28) for the case of natural broadening

† In fact $\bar{N} N_v$ represents the population inversion density in the cavity medium, i e , it is equal to $(N_2 - N_1)$ of Chapter 4 [see discussion after Eq (5 3-27)]

The minimum value of threshold inversion would correspond to $\omega_n = \omega_{21}$ (i e , at resonance) giving†

$$\bar{N}_{tm} = \frac{\epsilon_0 \hbar \gamma_{12}}{\mathcal{P}^2 Q'_n N_v} \tag{5 3-47}$$

Next, we substitute for χ'_n from Eq (5 3-40) in Eq (5 2-39) to obtain

$$\omega_n - \Omega_n = \frac{\omega_n}{2\hbar\epsilon_0}\mathcal{P}^2 \bar{N}N_v \frac{\omega_{21} - \omega_n}{(\omega_n - \omega_{21})^2 + \gamma_{12}^2} \tag{5 3-48}$$

where we have neglected the term ϕ_n in Eq (5 2-39) Thus, in the presence of the active medium, the oscillations do not occur at the passive cavity resonances but are shifted because of the presence of the χ'_n term In general, this shift is small and one can obtain the approximate oscillation frequencies as

$$\omega_n \approx \Omega_n + \frac{\Omega_n \mathcal{P}^2}{2\epsilon_0 \hbar} \bar{N}N_v \frac{\omega_{21} - \Omega_n}{(\Omega_n - \omega_{21})^2 + \gamma_{12}^2} \tag{5 3-49}$$

If Ω_n coincides exactly with the resonance frequency ω_{21} then $\omega_n = \Omega_n$ and in such a case the frequency of oscillation in the active resonator is the same as in the passive case If $\Omega_n < \omega_{21}$, then for an inverted medium, $\omega_n > \Omega_n$ Similarly for $\Omega_n > \omega_{21}$, $\omega_n < \Omega_n$ Thus, in the presence of the active medium, the oscillation frequencies are pulled towards the line center

At threshold, we substitute for $\bar{N}_t$ from Eq (5 3-44) to obtain

$$\omega_n - \Omega_n \approx \frac{\omega_n}{2Q'_n \gamma_{12}}(\omega_{21} - \omega_n) \tag{5 3-50}$$

where we have assumed $\omega_n \approx \omega_{21}$ We define a parameter

$$S = \frac{\omega_n / 2Q'_n}{\gamma_{12}} \tag{5 3-51}$$

which is known as the stabilizing factor‡, so that

$$\omega_n - \Omega_n \approx S(\omega_{21} - \omega_n)$$

or

$$\omega_n \approx \frac{\Omega_n + S\omega_{12}}{1 + S} \tag{5 3-52}$$

† Notice that for $\omega_n \neq \omega_{21}$, i e , for a mode shifted away from resonance, the value of N_t increases with increase in the value of $|\omega_n - \omega_{21}|$

‡ It represents the ratio of the cavity bandwidth to the natural linewidth

For a gas laser $S \sim 0.01$–0.1 so that the oscillation frequency lies very close to the normal mode frequency of the passive cavity mode

5.3.2 Higher-Order Theory

We have shown earlier that if the laser operates above threshold [see Eq (5.3-43)], the power will grow exponentially with time [see Eq (5.3-42)]. This unlimited growth is due to the assumption that the population difference remains constant with time [see Eq (5.3-27)]. However, as the power increases, the population of the upper level would decrease (because of increase in stimulated emission), and hence in an actual laser, the power level would saturate at a certain value. We will show this explicitly in this section†

Similar to the rate equations discussed in Chapter 4, we start with the equations describing the population of the two levels

$$\dot{\rho}_{11} = \lambda_1 - \gamma_1\rho_{11} + \left(\frac{i}{\hbar} H'_{21}\rho_{12} + \text{c c}\right) \qquad (5\ 2\text{-}53\text{a})$$

$$\dot{\rho}_{22} = \lambda_2 - \gamma_2\rho_{22} - \left(\frac{i}{\hbar} H'_{21}\rho_{12} + \text{c c}\right) \qquad (5\ 3\text{-}53\text{b})$$

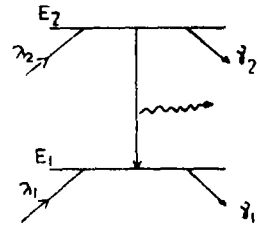



Fig 5.3 λ_1 and λ_2 represent the rates of pumping of the lower and upper levels respectively and γ_1 and γ_2 represent their decay constants

Just as in Section 4.4, the quantities λ_1 and λ_2 represent constant rates of pumping of atoms into levels 1 and 2, respectively (see Fig 5.3). In order to solve the above equations, we substitute for ρ_{12} from the first-order solution obtained in the previous section. Thus

$$\frac{i}{\hbar} H'_{21}\rho_{12} = \frac{i}{\hbar}[-E_n \mathscr{P} \cos(\omega_n t + \phi_n)\sin K_n z]\frac{\mathscr{P}}{2\hbar} N(z)\frac{\sin K_n z}{\Gamma_n} E_n(t)e^{i\theta} e^{i(\omega_n t+\phi_n)}$$

$$= -\frac{i}{2\hbar^2}\mathscr{P}^2 \frac{E_n^2(t)}{\Gamma_n}(\rho_{22} - \rho_{11})\sin^2 K_n z \cos(\omega_n t + \phi_n)e^{i(\omega_n t+\phi_n+\theta_n)}$$

† See also Section 4.4 where we showed that on a steady state basis the inversion can never exceed the threshold value

Hence

$$\frac{i}{\hbar} H'_{21}\rho_{12} + \text{c c} = \frac{\mathscr{P}^2}{2\hbar^2}\frac{E_n^2(t)}{\Gamma_n} \sin^2 K_n z \,(\rho_{22} - \rho_{11})G \qquad (5\ 3\text{-}54)$$

where

$$\begin{aligned} G &= -\cos(\omega_n t + \phi_n)[ie^{i(\omega_n t + \phi_n + \theta_n)} - ie^{-i(\omega_n t + \phi_n + \theta_n)}] \\ &= 2\cos(\omega_n t + \phi_n)\sin(\omega_n t + \phi_n + \theta_n) \\ &= 2\cos^2(\omega_n t + \phi_n)\sin\theta_n + \sin 2(\omega_n t + \phi_n)\cos\theta_n \\ &\simeq \sin\theta_n = \frac{\gamma_{12}}{\Gamma_n} \end{aligned} \qquad (5\ 3\text{-}55)$$

where we have replaced G by its time average value Substituting in Eq (5 3-53a), we get

$$\rho_{11} = \lambda_1 - \gamma_1\rho_{11} + R(\rho_{22} - \rho_{11}) \qquad (5\ 3\text{-}56)$$

Similarly

$$\rho_{22} = \lambda_2 - \gamma_2\rho_{22} - R(\rho_{22} - \rho_{11}) \qquad (5\ 3\text{-}57)$$

where

$$R = \frac{\gamma_{12}\mathscr{P}^2}{2\hbar^2} E_n^2 \frac{\sin^2 K_n z}{\Gamma_n^2} \qquad (5\ 3\text{-}58)$$

At steady state we must have $\rho_{11} = \rho_{22} = 0$, and

$$\rho_{11} - \frac{R}{\gamma_1}(\rho_{22} - \rho_{11}) = \frac{\lambda_1}{\gamma_1} \qquad (5\ 3\text{-}59\text{a})$$

$$\rho_{22} + \frac{R}{\gamma_2}(\rho_{22} - \rho_{11}) = \frac{\lambda_2}{\gamma_2} \qquad (5\ 3\text{-}59\text{b})$$

or

$$\rho_{22} - \rho_{11} = \frac{N(z)}{1 + R/R_s} \qquad (5\ 3\text{-}60)$$

where

$$N(z) = \frac{\lambda_2}{\gamma_2} - \frac{\lambda_1}{\gamma_1} \qquad (5\ 3\text{-}61)$$

and

$$R_s = \left(\frac{1}{\gamma_1} + \frac{1}{\gamma_2}\right)^{-1} = \frac{\gamma_1\gamma_2}{2\gamma_{12}} \qquad (5\ 3\text{-}62)$$

It follows from Eq (5 3-60) that the population inversion depends on the field value also In the absence of the field, $R = 0$ and the population difference density is simply $N(z)N_v$, however, as the field strength E_n (and hence R) increases, the population difference decreases Since $R(z)$ has a sinusoidal dependence on z [see Eq (5 3-58)], the population difference also varies with z Whenever $K_n z$ is an odd multiple of $\pi/2$ [i e , wherever the field has a maximum amplitude—see Eq (5 3-24)], the population difference has a minimum value, which is often referred to as hole burning in the population difference and the holes have a spacing of half of a wavelength

If instead of Eq (5 3-27), we now use Eq (5 3-60) for $\rho_{22} - \rho_{11}$, Eq (5 3-33) would be replaced by

$$\rho_{21} \simeq \frac{\mathcal{P}}{2\hbar}\frac{E_n \sin K_n z}{\Gamma_n}\frac{N(z)}{1 + R/R_s} e^{-i\theta_n} e^{-i(\omega_n t + \phi_n)} \tag{5 3-63}$$

Thus [cf Eq (5 3-38)]

$$p_n(t) = \frac{\mathcal{P}^2 E_n(t) e^{-i\theta_n}}{\hbar\Gamma_n} N_v \left[\frac{2}{L}\int_0^L \frac{N(z)}{1 + R/R_s} \sin^2 K_n z \, dz\right] \tag{5 3-64}$$

We next assume E_n (and hence R) to be small enough so that

$$\left(1 + \frac{R}{R_s}\right)^{-1} \simeq 1 - \frac{R}{R_s} \tag{5 3-65}$$

Substituting this in Eq (5 3-64) and carrying out a term-by-term integration, we obtain

$$\begin{aligned} \frac{2}{L}\int_0^L \frac{N(z)}{1 + R/R_s} \sin^2 K_n z \, dz &\simeq \frac{2}{L}\int_0^L N(z) \sin^2 K_n z \, dz \\ &\quad - \frac{\mathcal{P}^2 E_n^2 \gamma_{12}^2}{\hbar^2 \gamma_1 \gamma_2} \frac{1}{\Gamma_n^2}\left[\frac{2}{L}\int_0^L N(z) \sin^4 K_n z \, dz\right] \\ &\simeq \bar{N}\left[1 - \frac{3}{4}\frac{\mathcal{P}^2 E_n^2}{\hbar^2 \gamma_1 \gamma_2}\frac{\gamma_{12}^2}{\Gamma_n^2}\right] \end{aligned} \tag{5 3-66}$$

where we have used Eq (5 3-39) and the relation

$$\begin{aligned} \frac{2}{L}\int_0^L N(z) \sin^4 K_n z \, dz &\simeq \frac{2}{L}\left[\int_0^L N(z) \, dz\right] \langle \sin^4 K_n z \rangle \\ &\simeq \tfrac{3}{4}\bar{N} \end{aligned} \tag{5 3-67}$$

Thus,

$$p_n(t) \simeq \frac{\mathcal{P}^2 E_n}{\hbar\Gamma_n} e^{-i\theta_n} \bar{N} N_v \left(1 - \frac{3}{4}\frac{\mathcal{P}^2 E_n^2}{\hbar^2 \gamma_1 \gamma_2}\frac{\gamma_{12}^2}{\Gamma_n^2}\right) \tag{5 3-68}$$

or

$$p_n(t) \simeq \frac{\mathscr{P}^2 E_n}{\hbar \Gamma_n} e^{-i\theta_n} \bar{N} N_v \left(1 + \frac{3}{4} \frac{\mathscr{P}^2 E_n^2}{\hbar^2 \gamma_1 \gamma_2} \frac{\gamma_{12}^2}{\Gamma_n^2}\right)^{-1} \tag{5.3-69}$$

where in the last step we have assumed the two terms inside the square brackets in Eq. (5.3-68) to be the first two terms of a geometric series. This way, the gain saturates as the electric field increases indefinitely. Equation (5.3-69) may be compared with Eq. (5.3-38), hence instead of Eqs. (5.3-40) and (5.3-41), we get

$$\chi_n' \simeq \frac{\mathscr{P}^2 \bar{N} N_v}{\hbar \epsilon_0 \Gamma_n} \cos\theta_n \left(1 + \frac{3}{4} \frac{\mathscr{P}^2 E_n^2}{\hbar^2 \gamma_1 \gamma_2} \frac{\gamma_{12}^2}{\Gamma_n^2}\right)^{-1} \tag{5.3-70}$$

$$\chi_n'' \simeq -\frac{\mathscr{P}^2 \bar{N} N_v}{\hbar \epsilon_0 \Gamma_n} \sin\theta_n \left(1 + \frac{3}{4} \frac{\mathscr{P}^2 E_n^2}{\hbar^2 \gamma_1 \gamma_2} \frac{\gamma_{12}^2}{\Gamma_n^2}\right)^{-1} \tag{5.3-71}$$

Substituting this value of χ_n'' in Eq. (5.2-40), we get at steady state,

$$0 = \dot{E}_n = \left[-\frac{1}{2}\frac{\omega_n}{Q_n'} + \frac{1}{2}\omega_n \frac{\mathscr{P}^2 \bar{N} N_v}{\hbar \epsilon_0 \Gamma_n} \sin\theta_n \left(1 + \frac{3}{4} \frac{\mathscr{P}^2 E_n^2}{\hbar^2 \gamma_1 \gamma_2} \frac{\gamma_{12}^2}{\Gamma_n^2}\right)^{-1}\right] E_n \tag{5.3-72}$$

which after some simplification gives

$$E_n^2 = \frac{4\hbar^2}{3\mathscr{P}^2} \gamma_1 \gamma_2 \left[\frac{\bar{N}}{\bar{N}_{tm}} - 1 - \frac{(\omega_n - \omega_{21})^2}{\gamma_{12}^2}\right] \tag{5.3-73}$$

where $\bar{N}_{tm}$ is given by Eq. (5.3-47). The above equation gives the dependence of the saturation value of the intensity as a function of the detuning $(\omega_n - \omega_{21})$. At resonance $\omega_n = \omega_{21}$ and we get

$$E_n^2 = \frac{4\hbar^2 \gamma_1 \gamma_2}{3\mathscr{P}^2} \left(\frac{\bar{N}}{\bar{N}_{tm}} - 1\right) \tag{5.3-74}$$

It is clear from the above equation that the intensity of the field inside the cavity increases linearly with the pumping rate above threshold.

It should be pointed out that in the above equation, $\bar{N}_{tm}$ is proportional to the pumping rate at threshold [see Eq. (5.3-61)] and happens to be equal to the inversion density at the threshold [see Eq. (5.3-27)]. On the other hand, $\bar{N}$ is proportional to the pumping rate corresponding to the actual laser operation which is greater than $\bar{N}_t$. Thus if we assume $\gamma_1 \ggg \gamma_2$, then $\bar{N} N_v$ is nothing but R/T_{21} of Section 4.4. Using this value of $\bar{N}$ one obtains

$$E_n^2 \simeq \frac{4}{3} \frac{\hbar^2 \gamma_1 \gamma_2}{\mathscr{P}^2} \left(\frac{R}{R_t} - 1\right) \tag{5.3-75}$$

where we have used the relation

$$R_t = N_{2t}T_{21} \simeq N_v \bar{N}_t \gamma_2 \tag{5 3-76}$$

Further, in order to relate the photon number of the cavity to E_n^2, we note that the energy density of the field in the cavity is given by $\frac{1}{2}\epsilon_0 E_n^2$ and the total energy in the cavity of volume V would be $\frac{1}{2}\epsilon_0 V E_n^2$. If the frequency of the cavity mode is ω_n, the number of photons in the cavity mode would be

$$n = \frac{1}{2}\epsilon_0 \frac{E_n^2 V}{\hbar\omega_n} \tag{5 3-77}$$

Substituting this in Eq (5 3-75) and using the fact that K defined in Eq (4 4-5) is identical to†

$$K = \frac{\wp^2 \omega_n}{\hbar\epsilon_0 \gamma_{12}} \tag{5 3-78}$$

we obtain

$$n = \frac{4}{3}V\left(\frac{\gamma_1\gamma_2}{2\gamma_{12}}\right)\frac{1}{K}\left(\frac{R}{R_t} - 1\right) \tag{5 3-79}$$

For $\gamma_1 \gg \gamma_2$ (i e , the lower level has a very short lifetime as compared to the upper state)

$$\frac{\gamma_1\gamma_2}{2\gamma_{12}} = \frac{\gamma_1\gamma_2}{\gamma_1 + \gamma_2} \simeq \gamma_2 \tag{5 3-80}$$

Thus Eq (5 3-79) becomes

$$n = \frac{4}{3}V\frac{\gamma_2}{K}\left(\frac{R}{R_t} - 1\right) \tag{5 3-81}$$

which is the same as Eq (4 4-13b) obtained in the last chapter (γ_2 correspond to T_{21} of Section 4 4), apart from the factor 4/3 which has appeared because of the consideration of the spatial dependence of the modal field in this chapter

† Use has been made of Eq (5 3-46)

6

Optical Resonators

6.1 Introduction

In Chapter 1 we discussed briefly the optical resonator, which consists of a pair of mirrors facing each other in between which is placed the active laser medium which provides the amplification. As we discussed, the mirrors provide an optical feedback and the system then acts as an oscillator. In the present chapter, we give a more detailed account of optical resonators. In Section 6.2 we discuss the modes of a rectangular cavity and show that there exists an extremely large number of modes of oscillation under the linewidth of the active medium in a closed cavity of practical dimensions (which are large compared to the wavelength of light). Section 6.3 discusses the important concept of the quality factor of an optical resonator. In this section we obtain the linewidth corresponding to the passive cavity in terms of the parameters of the resonator. We also introduce the concept of cavity lifetime. In Section 6.4 we discuss the ultimate linewidth of a laser oscillator—this is, as discussed earlier in Chapter 5, determined by the spontaneous emissions occurring in the cavity. In practice, the observed linewidth is much larger than this ultimate width and is determined by mechanical stability, temperature fluctuations, etc. Section 6.5 discusses some techniques for selecting a single transverse and longitudinal mode in a laser oscillator. In Sections 6.6 and 6.7 we discuss the techniques for producing short intense pulses of light using Q switching and mode locking. Using the mode-locking techniques one can indeed obtain ultrashort pulses of very high peak power which find widespread applications.

In Section 6.8, we give a scalar wave analysis of the modes of a symmetric confocal resonator which consists of a pair of concave mirrors of equal radii of curvatures and separated by a distance equal to the radius of curvature. We will show that in such a structure, the lowest-order transverse mode has a Gaussian field distribution across its wave front. Most practical lasers are made to oscillate in this mode. In Section

6.9, we give the analysis and some results of the numerical analysis performed by Fox and Li (1961) for a planar resonator. In Section 6.10 we give the results for the beam width and the field distributions corresponding to a general spherical resonator. Section 6.11 discusses the geometrical optical analysis of general spherical resonators which are bounded by a pair of spherical mirrors of arbitrary radii of curvatures and separated by an arbitrary distance. We will obtain the conditions to be satisfied by the radii of curvatures and the separation between the mirrors so that the resonator is stable.

6.2 *Modes of a Rectangular Cavity and the Open Planar Resonator*

Consider a rectangular cavity of dimensions $2a \times 2b \times d$ as shown in Fig. 6.1. Starting from Maxwell's equations [see Eqs. (8.2-1)–(8.2-4)] one can show that the electric and magnetic fields satisfy a wave equation of the form given by

$$\nabla^2 \mathscr{E} - \frac{1}{c^2}\frac{\partial^2 \mathscr{E}}{\partial t^2} = 0 \qquad (6.2\text{-}1)$$

where c represents the velocity of light in the medium filling the rectangular cavity which is assumed here to be free space. Equation (6.2-1) can be obtained by taking the curl of Eq. (8.2-2), giving

$$\nabla \times (\nabla \times \mathscr{E}) = -\mu_0 \frac{\partial}{\partial t}(\nabla \times \mathscr{H}) = -\epsilon_0\mu_0 \frac{\partial^2 \mathscr{E}}{\partial t^2}$$

If we now use the identity

$$\nabla \times (\nabla \times \mathscr{E}) = \nabla(\nabla \cdot \mathscr{E}) - \nabla^2\mathscr{E} = -\nabla^2\mathscr{E}$$

[where we have used Eq. (8.2-3)], we immediately obtain Eq. (6.2-1).

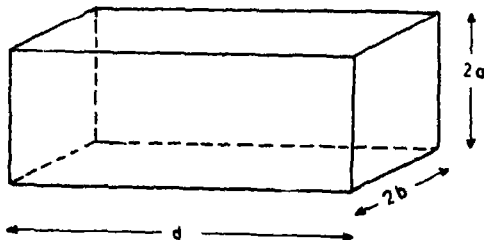



Fig. 6.1. A rectangular cavity of dimensions $2a \times 2b \times d$.

If the walls of the rectangular cavity are assumed to be perfectly conducting then the tangential component of the electric field must vanish at the walls Thus if $\hat{n}$ represents the unit vector along the normal to the wall then we must have

$$\mathscr{E} \times \hat{n} = 0 \tag{6 2-2}$$

on the walls of the cavity

Let us consider a Cartesian component (say the x component) of the electric vector this will also satisfy the wave equation which in the Cartesian system of coordinates will be given by

$$\frac{\partial^2 \mathscr{E}_x}{\partial x^2} + \frac{\partial^2 \mathscr{E}_x}{\partial y^2} + \frac{\partial^2 \mathscr{E}_x}{\partial z^2} = \frac{1}{c^2}\frac{\partial^2 \mathscr{E}_x}{\partial t^2} \tag{6 2-3}$$

In order to solve Eq (6 2-3) we use the method of separation of variables and write

$$\mathscr{E}_x = X(x)Y(y)Z(z)T(t) \tag{6 2-4}$$

Substituting this in Eq (6 2-3) and dividing by $\mathscr{E}_x$ we obtain

$$\frac{1}{X}\frac{d^2X}{dx^2} + \frac{1}{Y}\frac{d^2Y}{dy^2} + \frac{1}{Z}\frac{d^2Z}{dz^2} = \frac{1}{c^2T}\frac{d^2T}{dt^2} \tag{6 2-5}$$

Thus the variables have indeed separated out and we may write

$$\frac{1}{X}\frac{d^2X}{dx^2} = -k_x^2 \tag{6 2-6}$$

$$\frac{1}{Y}\frac{d^2Y}{dy^2} = -k_y^2 \tag{6 2-7}$$

$$\frac{1}{Z}\frac{d^2Z}{dz^2} = -k_z^2 \tag{6 2-8}$$

and

$$\frac{1}{c^2T}\frac{d^2T}{dt^2} = -k^2 \tag{6 2-9}$$

where

$$k^2 = k_x^2 + k_y^2 + k_z^2 \tag{6 2-10}$$

Equation (6 2-9) tells us that the time dependence is of the form

$$T(t) = Ae^{-i\omega t} \tag{6 2-11}$$

where $\omega = ck$ represents the angular frequency of the wave and A is a

constant It should be mentioned that we could equally well have chosen the time dependence to be of the form $e^{i\omega t}$ Since $\mathcal{E}_x$ is a tangential component on the planes $y=0$, $y=2b$, $z=0$, and $z=d$, it has to vanish on these planes and the solutions of Eqs (6 2-7) and (6 2-8) would be $\sin k_y y$ and $\sin k_z z$, respectively, with

$$k_y = \frac{n\pi}{2b}, \qquad k_z = \frac{q\pi}{d}, \qquad n, q = 0, 1, 2, 3, \tag{6 2-12}$$

where we have intentionally included the value 0, which in this case would lead to the trivial solution of $\mathcal{E}_x$ vanishing everywhere In a similar manner, the x and z dependences of $\mathcal{E}_y$ would be $\sin k_x x$ and $\sin k_z z$, respectively, with

$$k_x = \frac{m\pi}{2a}, \qquad m = 0, 1, 2, \tag{6 2-13}$$

and k_z given by Eq (6 2-12) Finally, the x and y dependences of $\mathcal{E}_z$ would be $\sin k_x x$ and $\sin k_y y$, respectively

Now, because of the above forms of the x dependence of $\mathcal{E}_y$ and $\mathcal{E}_z$, $\partial\mathcal{E}_y/\partial y$ and $\partial\mathcal{E}_z/\partial z$ would vanish on the surfaces $x=0$ and $x=2a$ Thus, on the planes $x = 0$ and $x = 2a$, the equation $\nabla \cdot \mathcal{E} = 0$ leads to $\partial\mathcal{E}_x/\partial x = 0$ Hence the x dependence of $\mathcal{E}_x$ will be of the form $\cos k_x x$, with k_x given by Eq (6 2-13) Notice that the case $m = 0$ now corresponds to a nontrivial solution

In a similar manner one may obtain the solutions for $\mathcal{E}_y$ and $\mathcal{E}_z$ The complete solution (apart from the time dependence) would therefore be given by

$$\begin{aligned} E_x &= E_{0x} \cos k_x x \sin k_y y \sin k_z z \\ E_y &= E_{0y} \sin k_x x \cos k_y y \sin k_z z \\ E_z &= E_{0z} \sin k_x x \sin k_y y \cos k_z z \end{aligned} \tag{6 2-14}$$

where E_{0x} E_{0y}, and E_{0z} are constants The use of the Maxwell equation $\nabla \cdot \mathbf{E} = 0$ immediately gives

$$\mathbf{E}_0 \cdot \mathbf{k} = 0 \tag{6 2-15}$$

where $\mathbf{k} = \hat{\mathbf{x}}k_x + \hat{\mathbf{y}}k_y + \hat{\mathbf{z}}k_z$ Using Eqs (6 2-12), (6 2-13), and (6 2-10) we obtain

$$\begin{aligned} \omega^2 = c^2k^2 &= c^2(k_x^2 + k_y^2 + k_z^2) \\ &= c^2\pi^2 \left(\frac{m^2}{4a^2} + \frac{n^2}{4b^2} + \frac{q^2}{d^2}\right) \end{aligned}$$

or

$$\omega = c\pi\left(\frac{m^2}{4a^2} + \frac{n^2}{4b^2} + \frac{q^2}{d^2}\right)^{1/2} \tag{6 2-16}$$

which gives us the allowed frequencies of oscillation of the field in the cavity Field configurations given by Eq (6 2-14) represent standing wave patterns in the cavity and are called the modes of oscillation of the cavity†

Using Eq (6 2-16) we can show (see Appendix D) that the number of modes per unit volume in a frequency interval from ν to $\nu + d\nu$ will be given by

$$p(\nu)\, d\nu = \frac{8\pi}{c^3}\nu^2\, d\nu \tag{6 2-17}$$

Thus the number of modes per unit volume (in a closed cavity) that fall within the linewidth of an atomic system with $d\nu \sim 3 \times 10^9$ Hz at $\nu = 3\times10^{14}$ Hz would be

$$\frac{8 \times 3\;14}{27 \times 10^{30}} 9\times10^{28} \times 3 \times 10^9 \simeq 2 \times 10^8\,\mathrm{cm}^{-3} \tag{6 2-18}$$

Thus, for cavities of practical dimensions, the total number of modes which lie within an atomic linewidth is extremely large Now, in a cavity supporting such a large number of modes, all these oscillating modes will draw energy from the atoms and the resulting emission would be far from monochromatic In order that there be only a very few oscillating modes within the linewidth, one has to have cavities of the dimension of the wavelength of the radiation which in the visible and infrared portions of the spectrum is impractical

This problem is overcome by the use of open resonators which, in the simplest case, consists of a pair of plane mirrors facing each other and the lateral portions of the resonator are open (see Fig 1 1) The field components of the modes of a closed rectangular cavity are given by Eq (6 2-14) If we write each of the cosine and sine terms in terms of exponentials, we obtain for example

$$\begin{aligned}\mathscr{E}_x &\sim (e^{ik_x x} + e^{-ik_x x})(e^{ik_y y} - e^{-ik_y y})(e^{ik_z z} - e^{-ik_z z})e^{-i\omega t}\\ &= e^{i(\mathbf{k}_1\cdot\mathbf{r}-\omega t)} - e^{i(\mathbf{k}_2\cdot\mathbf{r}-\omega t)} + \cdots + e^{i(\mathbf{k}_8\cdot\mathbf{r}-\omega t)}\end{aligned}$$

† It may be noted that for given values of m, n and q two of the components of $\mathbf{E}$ determine the third through Eq (6 2-15) Thus a given mode will have two independent directions of polarization

where

$$\mathbf{k}_1 = k_x\hat{\mathbf{x}} + k_y\hat{\mathbf{y}} + k_z\hat{\mathbf{z}}$$
$$\mathbf{k}_2 = k_x\hat{\mathbf{x}} + k_y\hat{\mathbf{y}} - k_z\hat{\mathbf{z}}$$

$$\mathbf{k}_8 = -k_x\hat{\mathbf{x}} - k_y\hat{\mathbf{y}} - k_z\hat{\mathbf{z}}$$

Thus we see that a mode is formed by eight plane waves with wave vector components $\pm k_x$, $\pm k_y$, $\pm k_z$ along the x, y, and z directions, respectively These component plane waves are traveling along directions which have direction cosines along the x, y, and z directions as $\pm m\lambda/4a$, $\pm n\lambda/4b$, and $\pm q\lambda/2d$, respectively In open resonators, the lateral surfaces of the resonator are removed and one simply has a pair of mirrors facing each other (and normal to the z axis) as shown in Fig 1 1 Those modes which are formed by superposition of waves which travel approximately at right angles to the mirrors will keep oscillating between the mirrors with not too large a loss per transit, while waves traveling obliquely will be lost after very few traversals It may be mentioned that in an actual resonator those waves which are traveling at right angles to the mirrors will have an amplitude which decreases away from the axis (see Sections 6 8 and 6 9) and hence their losses (due to the finite size of the mirrors) will be smaller than when they have a uniform distribution of amplitude over the mirror surfaces

In order to get an approximate expression for the oscillation frequencies of the modes of the open resonator system, we observe that the modes of the open cavity will be approximately the modes of the closed cavity for which $m, n \ll q$, i e , will consist of those waves which are traveling at very small angles with the z-axis If we use the approximation m $n \ll q$ we may make a binomial expansion of Eq (6 2-16) and obtain

$$\omega_{mnq} \simeq c\pi\left[\frac{q}{d} + \left(\frac{m^2}{a^2} + \frac{n^2}{b^2}\right)\frac{d}{8q}\right] \tag{6 2-19}$$

For a pair of square mirrors, $a = b$ and we may write for the resonant frequencies

$$\nu_{mnq} = \frac{\omega_{mnq}}{2\pi} \simeq \frac{c}{2}\left[\frac{q}{d} + \frac{(m^2+n^2)}{q}\frac{d}{8a^2}\right] \tag{6 2-20}$$

The difference in frequency of oscillation between two modes having the same values of m and n but differing in q by unity is very nearly given by

$$\Delta\nu_q \simeq \frac{c}{2d} \tag{6 2-21}$$

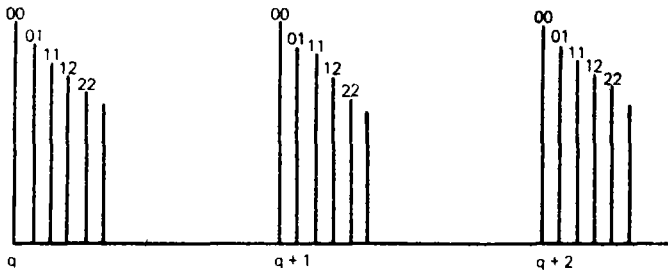



Fig 6 2 The resonance frequencies of a plane parallel resonator Notice that the spacing between two modes differing in transverse mode number by unity is much less than the spacing between two longitudinal modes The numbers on the line represent the values of m and n

Since q specifies the field distribution along the z axis [see Eq (6 2-12)], modes differing in their q values are referred to as various longitudinal modes For a typical laser resonator, $d \simeq 100$ cm and one obtains a longitudinal mode spacing $\sim 1.5 \times 10^8$ Hz or 150 MHz

Modes differing in their m and n values have different field distributions along the transverse direction and are referred to as the transverse modes of the resonator The frequency difference between two consecutive transverse modes having the same q value and differing in m by unity would be

$$\Delta\nu_m = \frac{c}{2}\frac{d}{4a^2 q}(m + \tfrac{1}{2})$$

$$\simeq \frac{cd}{2a^2}\frac{\lambda}{2d}\tfrac{1}{4}(m + \tfrac{1}{2}) = \frac{c\lambda}{16a^2}(m + \tfrac{1}{2})$$

$$= \Delta\nu_q \frac{\lambda d}{8a^2}(m + \tfrac{1}{2}) \qquad (6\,2\text{-}22)$$

where we have used the fact that† $q \simeq 2d/\lambda$ For typical resonators, $\lambda \sim 8 \times 10^{-5}$ cm, $d = 100$ cm, $a = 1$ cm $\lambda d/8a^2 \simeq 10^{-3}$ Since the value of m is of the order of unity $\Delta\nu_m \ll \Delta\nu_q$ Thus for such a resonator the frequency difference between modes differing in transverse mode number is much less than the frequency separation between two longitudinal modes (see Fig 6 2)

In Section 6 9 we present a more rigorous scalar analysis of the modes of a resonator consisting of two plane parallel mirrors facing each other this analysis is based on the formulation of Fox and Li (1961) The

† This follows from the fact that since the waves constituting the mode are approximately traveling along the z direction $k \simeq k_z \simeq 2\pi/\lambda$ and using Eq (6 2 12) we get $q \simeq 2d/\lambda$

modes are obtained by a numerical technique and it will be seen that the lowest-order transverse mode has a large amplitude at the center of the mirror and the amplitude decreases away from the center Thus the losses due to diffraction are much smaller than for a uniform plane wave

A pair of curved mirrors facing each other can, in general form a resonant cavity But there exist a class of resonators referred to as stable resonators in which the electromagnetic field can keep bouncing back and forth in between the mirrors without much loss due to the finite size of the mirrors On the other hand, in unstable resonator configurations, the field escapes from the sides of the mirror and is not well confined in the cavity An interesting example of a stable resonator system is the symmetric confocal resonator system, which consists of a pair of concave mirrors of radii of curvatures R and separated by a distance equal to R so that the foci of both mirrors coincide at the center—hence the name confocal (see Fig 6 3) A detailed scalar wave analysis of such a system is discussed in Section 6 8 where it is shown that the modes of such a resonator system are the Hermite–Gauss functions, the lowest-order mode of such a resonator system has a Gaussian amplitude distribution across its wave front In addition to the confocal resonator system, there exist a large number of other resonator systems which are stable In Section 6 10 we show that a resonator made up of a pair of spherical mirrors can form a stable resonant cavity under certain conditions The lowest-order mode, in general, has a Gaussian field distribution with the beam parameters depending on the radii of curvatures of the mirrors and their separation and the higher-order modes are again Hermite–Gaussian Figure 6 4 shows the photographs of some of the lower-order modes of a stable resonator cavity It may be observed that higher-order modes have more spread-out distributions

In Section 6 11 we have given a geometrical optics analysis of a general spherical resonator In the language of geometrical optics, in stable resonators, a ray of light may keep bouncing back and forth between the mirrors indefinitely without ever escaping from the system

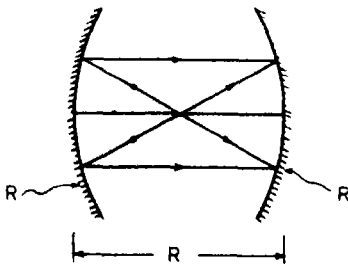



Fig 6 3 A symmetric confocal resonator system

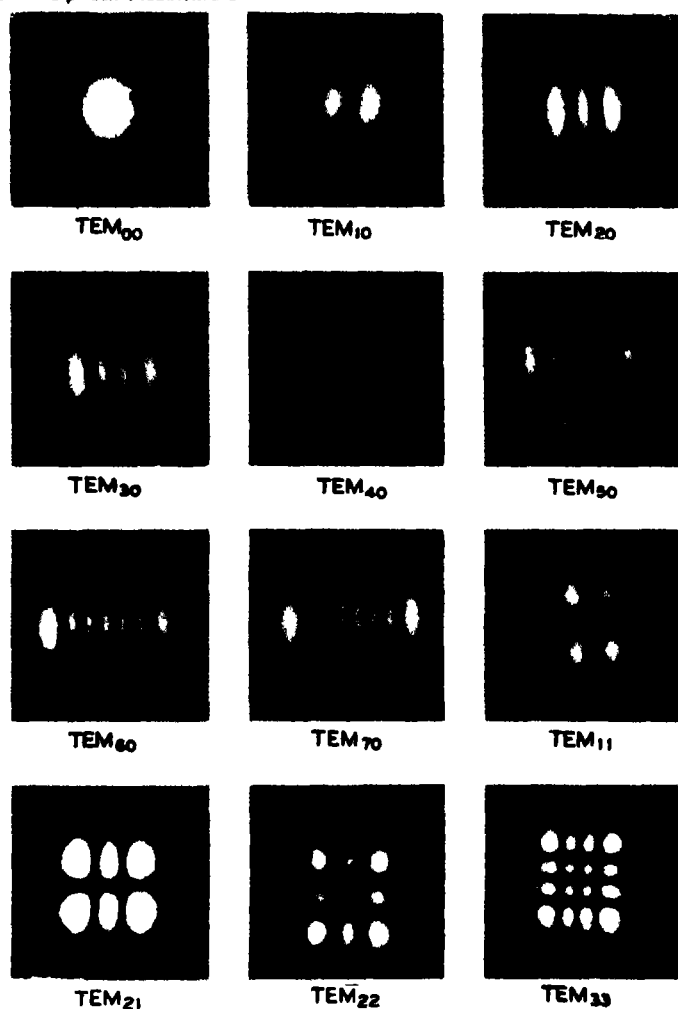



Fig 6 4 Photograph showing some of the lower-order resonator modes (After Kogelnik and Li 1966 Photograph courtesy of Dr H Kogelnik)

On the other hand, in an unstable resonator system, the ray diverges away from the axis and escapes from the resonator after a few traversals

6.3. The Quality Factor

Since the mirrors of a laser system form an open resonator cavity, there are always losses associated with any mode The main types of loss

mechanisms are the finite reflectivities of the mirrors (with the remaining energy partially absorbed and partially transmitted as the useful laser beam), scattering and absorption in the medium filling the resonator cavity and the diffraction spill-over when the field undergoes reflection from the mirrors This dissipation in energy is described in terms of the quality factor Q of the mode, which may be defined as

$$Q = \frac{\omega_0 \times \text{energy stored in the mode}}{\text{energy dissipated per second in that mode}} \tag{6 3-1}$$

where ω_0 $(=2\pi\nu_0)$ corresponds to the oscillation frequency of a mode Thus if $W(t)$ represents the energy stored in the mode at time t, then from the definition of Q,

$$\frac{dW}{dt} = -\frac{\omega_0}{Q} W \tag{6 3-2}$$

the negative sign takes care of the fact that there is usually a loss in energy One can easily integrate Eq (6 3-2) and obtain

$$W(t) = W(t = 0)e^{-\omega_0 t/Q} \tag{6 3-3}$$

This corresponds to an exponential decay of energy with time It follows from Eq (6 3-3) that the energy decays to $1/e$ of its value at $t = 0$ in a time

$$t_c = \frac{Q}{\omega_0} = \frac{Q}{2\pi\nu_0} \tag{6 3-4}$$

which is also referred to as the passive cavity lifetime Thus for a mode we can associate a field given by

$$E(t) = E_0 e^{-\omega_0 t/2Q} e^{2\pi i \nu_0 t} \tag{6 3-5}$$

The frequency spectrum associated with this wave train (which extends from $t = 0$ to $t = \infty$) is obtained by taking the Fourier transform of Eq (6 3-5) (see Appendix A)

$$\begin{aligned} \tilde{E}(\nu) &= \int_0^\infty E(t)e^{-2\pi i \nu t}\, dt \\ &= \frac{E_0}{2\pi[i(\nu - \nu_0) + \nu_0/2Q]} \end{aligned} \tag{6 3-6}$$

The frequency distribution of the intensity would be given by

$$\begin{aligned} I(\nu) &= |\tilde{E}(\nu)|^2 \\ &= \frac{E_0^2}{4\pi^2[(\nu - \nu_0)^2 + \nu_0^2/4Q^2]} \end{aligned} \tag{6 3-7}$$

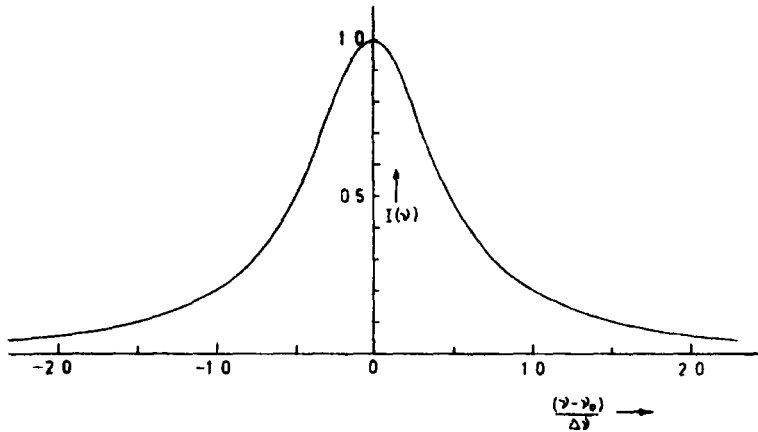



Fig 6 5 A Lorentzian distribution given by Eq (6 3-7)

The frequency dependence characterized by Eq (6 3-7) is peaked at $\nu = \nu_0$ and corresponds to a Lorentzian distribution (see Fig 6 5) As can be seen from Eq (6 3-7), the intensity decreases to half the peak value at $\nu = \nu_0 + \nu_0/2Q$ and $\nu = \nu_0 - \nu_0/2Q$ Thus the full width at half-maximum (FWHM) is

$$\Delta\nu = \frac{\nu_0}{Q} \tag{6 3-8}$$

From Eq (6 3 8) it is clear that the width of the output spectrum depends inversely on the quality factor Q associated with that mode, the smaller the losses in a mode, the higher is the value of Q and hence the smaller is the frequency half-width $\Delta\nu$

In order to calculate the Q of a passive resonator, we first find the energy left in the cavity after one complete cycle of oscillation and then use Eq (6 3-3) to obtain an explicit expression for Q Let W_0 be the total energy contained within the cavity at $t = 0$ One complete cycle of oscillation in the cavity corresponds to a pair of reflections from mirrors M_1 and M_2 (with power reflection coefficients R_1 and R_2, respectively) and two traversals through the laser medium, which is assumed to have a net power absorption coefficient α_c per unit length Thus the energy remaining within the cavity after one complete cycle would be

$$W_0 R_1 R_2 e^{-2\alpha d} \tag{6 3-9}$$

where d is the length of the cavity Also, one complete cycle corresponds to

$$t = \frac{2n_0 d}{c} \tag{6 3-10}$$

where n_0 is the refractive index of the medium inside the cavity Thus from Eq (6 3-3), we obtain for the energy inside the cavity

$$W_0 \exp\left(-\frac{2\pi\nu_0}{Q}\frac{2n_0 d}{c}\right) \tag{6 3-11}$$

Comparing Eqs (6 3-9) and (6 3-11) we obtain

$$2\alpha_c d - \ln R_1 R_2 = \frac{4\pi\nu_0 n_0 d}{Qc}$$

or

$$Q = \frac{4\pi n_0 \nu_0 d}{c} \frac{1}{2\alpha_c d - \ln R_1 R_2} \tag{6 3-12}$$

We can also obtain an expression for the cavity lifetime t_c in terms of the fractional loss per round trip The initial energy W_0 becomes $W_0 \exp(-\kappa)$ after one round trip, here

$$\kappa = 2\alpha_c d - \ln R_1 R_2 \tag{6 3-13}$$

Thus, the fractional loss per round trip would be

$$x = \frac{W_0 - W_0 e^{-\kappa}}{W_0} = 1 - e^{-\kappa}$$

or

$$\kappa = \ln\left(\frac{1}{1-x}\right) \tag{6 3-14}$$

Hence from Eqs (6 3-4) and (6 3-12)–(6 3-14) we get

$$t_c = \frac{2n_0 d}{c \ln[1/(1-x)]} = \frac{2n_0 d}{c(2\alpha_c d - \ln R_1 R_2)} \tag{6 3-15}$$

From Eq (6 3-8) we may write

$$\Delta\nu_p = \frac{c}{2\pi n_0 d}[\alpha_c d - \tfrac{1}{2}\ln R_1 R_2] \tag{6 3-16}$$

where the subscript p stands for passive cavity For a typical cavity $d = 100$ cm, $\alpha_c d - \frac{1}{2}\ln R_1 R_2 \sim 2 \times 10^{-2}$, and assuming $n_0 \simeq 1$, we obtain

$$\Delta\nu_p \simeq 1 \text{ MHz} \tag{6 3-17}$$

Observe that this width is much smaller than the separation between two adjacent longitudinal modes which for a cavity of length 100 cm is ~150 MHz (see Section 6 2)

6.4. The Ultimate Linewidth of the Laser

As we had seen in Section 6.3, the quality factor and hence the linewidth of oscillation is determined by the losses suffered by the energy in a particular mode. When the laser is oscillating in a steady state, the losses are exactly compensated by the gain due to stimulated and spontaneous emissions. Even though stimulated emission is responsible for compensating almost all the losses, there always exists a finite amount of spontaneous emission into every cavity mode of the system (see Section 8.5). The emission due to spontaneous transitions is incoherent with the field already existing in a particular mode, and it is this portion that determines the ultimate linewidth of the laser.

In order to obtain an expression for the ultimate linewidth, we recall Eq. (6.3-1), which defines the quality factor associated with a mode. The linewidth of oscillation is related to the quality factor through Eq. (6.3-8). Since the emission arising out of spontaneous transitions is incoherent, we regard this portion of the energy in the mode as a dissipation and it is this dissipation which determines the ultimate linewidth. We see from Eq. (4.4-7) that the rate of emission of photons due to spontaneous transition into a mode is KN_2. Since at and beyond threshold of oscillation the inversion does not exceed the value $1/Kt_c$ [where t_c is the passive cavity lifetime—see Eq. (4.4-8)], we may write for the rate of spontaneous emission into the mode for an oscillating laser $K(1/Kt_c) = 1/t_c$. Hence the amount of energy appearing in a mode (per unit time) due to spontaneous emission only is $h\nu_0/t_c$, where ν_0 represents the frequency of the oscillating mode. If there are n photons in the oscillating mode then, assuming that the cavity losses are only due to output coupling, we get for the output power of the laser

$$P^o = \frac{nh\nu_0}{t_c} \tag{6.4-1}$$

Thus we get the number of photons n present in the cavity mode in terms of the output power in that mode as

$$n = \frac{t_c}{h\nu_0} P^o \tag{6.4-2}$$

and the energy stored in the mode is $nh\nu_0 = t_c P^o$. Hence from Eqs. (6.3-1) and (6.3-8) we obtain

$$\begin{aligned} Q = \frac{\nu_0}{\delta\nu} &= \omega_0 \times \frac{\text{energy stored}}{\text{coherent energy dissipated per unit time}} \\ &= \omega_0 \times \frac{\text{energy stored}}{\text{energy arising out of spontaneous emissions per unit time}} \\ &= \frac{(2\pi\nu_0)t_c P^o}{(h\nu_0/t_c)} = \frac{2\pi t_c^2 P^o}{h} \end{aligned} \tag{6.4-3}$$

where $\delta\nu$ represents the linewidth of the output laser beam when the laser is oscillating in steady state The cavity lifetime t_c is given by the ratio of the passive cavity Q and $2\pi\nu_0$ Since Q is given by Eq (6 3-8) we get

$$t_c = \frac{Q}{2\pi\nu_0} = \frac{1}{2\pi(\Delta\nu_p)} \tag{6 4-4}$$

where $(\Delta\nu_p)$ represents the passive cavity linewidth [see Eq (6 3-16)] Thus we obtain from Eqs (6 4-3) and (6 4-4),

$$\delta\nu \simeq \frac{2\pi(\Delta\nu_p)^2 h\nu_0}{P^o} \tag{6 4-5}$$

The above equation represents the ultimate linewidth of a laser oscillator, and is similar to that as given by Schawlow and Townes (1958) It follows from Eq (6 4-5) that the linewidth $\delta\nu$ decreases with increase in power output Physically this is due to the fact that for a given mirror transmittance, increase in P^o corresponds to increase in the photon number in the cavity, which in turn implies a greater dominance of stimulated transitions over spontaneous transitions

The above treatment, which is a rather heuristic derivation of the ultimate linewidth, is similar to the one given by Maitland and Dunn (1969) and is based on an analysis by Gordon *et al* (1955) Recently Jacobs (1979) has considered the radiation field inside an oscillating laser cavity in a single mode to consist of n photons all corresponding to in-phase waves, except for spontaneous emission into the mode that occurs at the same frequency with a random phase Assuming that the ultimate linewidth is due to the phase jitter of the spontaneous photons, he has estimated the ultimate linewidth, which is half of that given by Eq (6 4-5)

We consider a typical example with $\Delta\nu_p = 1$ MHz, $P^o = 10^{-3}$ W, $h\nu = 2 \times 10^{-19}$ J (which corresponds to the red region of the spectrum) and one has

$$\delta\nu \simeq \frac{2 \times 3.14 \times 10^{12} \times 2 \times 10^{-19}}{10^{-3}} \simeq 10^{-3}\ \text{Hz} \tag{6 4-6}$$

which is indeed an extremely small quantity However, such small linewidths are never obtained in practice and the linewidth of the output laser beam is limited by the oscillations in the length of the cavity, temperature fluctuations, etc For example, if we assume $\nu \simeq qc/2d$ [see Eq (6 2-20)] then a change of Δd in d introduces a change in ν given by

$$\Delta\nu = -\frac{qc}{2d^2}\Delta d = \frac{\Delta d}{d}\nu$$

where we have disregarded the negative sign as it is of no significance here Thus, for maintaining a monochromaticity given by Eq (6 4-6), the cavity must be kept stable to an accuracy of

$$\Delta d \simeq \frac{10^2 \times 10^{-3}}{3 \times 10^{14}} \simeq 3 \times 10^{-16}\,\text{cm} = 3 \times 10^{-8}\,\text{Å}$$

which corresponds to a stability less than even nuclear dimensions!

6.5. Transverse and Longitudinal Mode Selection

Since optical resonators have dimensions which are large in comparison to the wavelength of light, there will, in general, be a large number of modes which fall within the atomic linewidth and which can oscillate in a laser Thus the output may consist of a number of longitudinal modes each oscillating at a different frequency Also, since the various transverse modes have different phase and amplitude distributions across the output plane, the resulting laser output may not be highly directional In fact the power per frequency and solid angle interval is maximum when the laser oscillates in a single mode which is the fundamental mode In order to obtain highly directional and spectrally pure laser output various techniques have been developed both for transverse and longitudinal mode selection in lasers In this section we will briefly discuss some of the important mode selection techniques

6 5 1 Transverse Mode Selection

As we show in Section 6 8, the lowest-order transverse mode has a Gaussian amplitude distribution across a transverse plane It is this mode that one usually prefers to work with as it has no abrupt phase changes across the wave front (as the higher-order transverse modes have) and has a monotonically decreasing amplitude away from the axis This leads to the fact that it can be focused to regions of the order of the wavelength of light, producing enormous intensities The Gaussian mode has the narrowest transverse distribution and hence if an aperture is introduced in the cavity which increases the loss of the next higher order transverse mode to a value where the loss exceeds the gain, the laser would oscillate only in the lowest-order mode Observe that this introduces additional losses even for the lowest-order mode Specific higher-order transverse modes can also be selected at a time by choosing complex apertures which introduce high loss for all modes except the required mode or by profiling the reflectivity of one of the mirrors to suit the required mode Thus, a

wire placed normal to the axis and passing through the resonator axis would select the second-order TE_{01} mode

6 5 2 Longitudinal Mode Selection

Even when a laser is oscillating in a single transverse mode, it could be oscillating in different longitudinal modes separated by a frequency spacing of $c/2d$ One possible method to obtain single-mode oscillation is to decrease the length of the cavity d to such a value that there is only one longitudinal mode within the gain profile which has its gain higher than the losses Such a method can be used with gas lasers which have relatively small linewidths It must be noted that as the length of the laser cavity is decreased, the volume of the active material also decreases and hence the output power would also decrease

In solid state lasers, the bandwidth is large and the above technique becomes impractical Hence various other techniques have been evolved for longitudinal mode selection For example, Fig 6 6 shows a Fabry–Perot etalon placed inside a laser resonator It can be shown that for an etalon of thickness t, the transmittance of the etalon has peaks at various frequencies given by (see, e g , Born and Wolf, 1975)

$$\nu = \frac{mc}{2n't\cos\theta'}, \qquad m = 1, 2, 3, \tag{6 5-1}$$

where n' is the refractive index of the etalon medium and θ' is the angle of refraction inside the etalon The frequency separation between two high-transmission (or low-loss modes in the composite resonator) would be $c/(2n't\cos\theta')$ Thus by choosing small values of t and by adjusting θ', only the mode at the center of the line may be made to oscillate

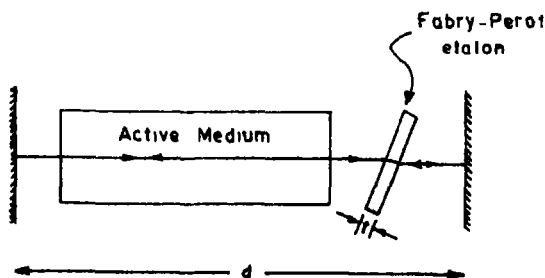



Fig 6 6 For selecting a single longitudinal mode one could place a Fabry–Perot etalon inside the cavity

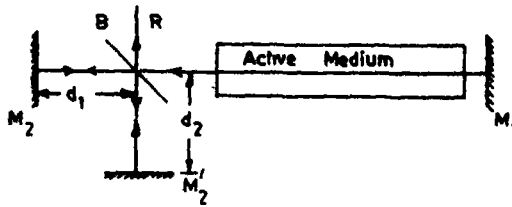



Fig 6 7 The Fox–Smith interferometer arrangement in which one of the resonator mirrors is replaced by a combination of a beam splitter and two mirrors M_2 and M_2' The resonator would oscillate only at those frequencies for which the amplitude of the reflected beam R is zero

Another scheme for longitudinal mode selection is the Fox–Smith interferometer arrangement depicted in Fig 6 7 It consists of a beam splitter B and two mirrors M_2 and M_2' (both being completely reflecting) forming the composite mirror on one end and another semitransparent mirror M_1 on the other end It can be shown that the reflectivity of the composite mirror arrangement has peaks which are spaced by a distance $c/[2(d_1 + d_2)]$ where d_1 and d_2 are the distances of the two mirrors M_2 and M_2' from B Thus by a proper choice one can achieve single longitudinal mode oscillation

In addition to the above two methods there exist many other techniques for longitudinal mode selection For further details readers are referred to Smith (1972)

6.6 Q Switching

The quality factor Q of a laser cavity is determined by the losses suffered by the modes of the cavity, smaller the losses higher the value of Q Consider a laser cavity in which a shutter is introduced in front of one mirror If the active medium is continuously pumped keeping the shutter closed, the population inversion in the cavity goes on increasing and reaches a high value If now the shutter is suddenly opened, the inversion would correspond to a value much above the threshold and the energy stored in the cavity will be released in the form of a short pulse of light with a high peak value of intensity Since the opening of the shutter increases the Q value from a very small value before opening to a very large value after opening, this technique of producing a short intense pulse of light is referred to as Q switching If the shutter is opened in a time much shorter than the time required for the building of laser oscillation, the output would consist of a giant pulse of light In case the

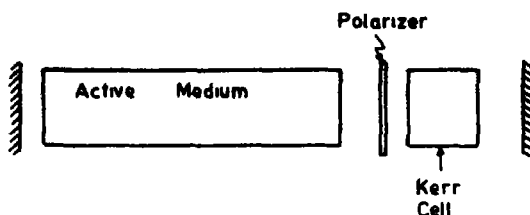



Fig 6 8 *Q* switching using a combination of a polarizer and a Kerr cell

shutter opening is slow, the output would be a series of pulses having smaller peak power

Various techniques have been developed for *Q* switching of lasers One method consists in a mechanical rotation of one of the laser mirrors about an axis normal to the resonator axis When the mirrors are not parallel, the losses in the resonator are large and the pump increases the inversion beyond the threshold corresponding to the case when the mirrors are parallel If the timing of the pump pulse is such that it reaches a maximum as the two mirrors are getting parallel, as soon as the mirrors become parallel, a giant pulse would appear at the output Since a mechanical switching is comparatively slow, one usually obtains pulse lengths of 25 to 50 nsec Typical rotation speeds are 30,000 revolutions per minute

In contrast to the mechanical switching, there exist other faster electronic switching techniques such as using the Kerr and Pockels effects† In such a method, one places a polarizer and the birefringent material in front of one of the mirrors (see Fig 6 8) The two are adjusted so that when the cell is based a double pass through the electro-optic material rotates the linearly polarized light through 90° and is blocked by the polarizer In such a position, the losses in the cavity would be large On removing the applied voltage on the cell, the cell loses its birefringence and the cell does not rotate the polarization In this position the losses are small and corresponds to an open shutter Typical values of voltage required for operation are a few kilovolts for a Pockels cell and a few tens of kilovolts for a Kerr cell

Q switching can also be obtained using saturable absorbers These absorbers have a transmittance which is constant at small incident powers, but at sufficiently high intensity, the transmittance begins to increase because of two-state saturation The saturation intensity (namely, the

† When the index of refraction of an electro-optic material changes linearly with the applied electric field it is termed the Pockels effect, while a quadratic dependence is termed the Kerr effect

intensity required to decrease the absorption coefficient to one half the low-power value) for normal dye solutions is about 10^7 W/cm^2 The operation of such a device may be understood as follows Owing to the small transmittance of the saturable absorber placed inside the cavity, laser oscillation cannot start until the attainment of a large population inversion As the power level inside the cavity goes on increasing, the dye begins to be bleached This bleaching results in a larger transmittance which in turn increases the power level inside the cavity The increased power results in a larger bleaching and thus the dye becomes almost transparent Since in this condition the inversion is much more than the threshold inversion, the gain is much more than the losses and thus a giant pulse is produced

6.7 Mode Locking in Lasers

In this section, we briefly discuss the technique of mode locking of lasers used for generating extremely short ($\sim 10^{-12}$ sec) optical pulses of high peak power In order to understand the concept of mode locking, we consider a laser formed by a pair of mirrors separated by a distance d enclosing the active medium with a linewidth $\Delta\nu$ about a central frequency ν_0 (see Fig 6 9) As we have discussed earlier, the frequency spacing of the longitudinal modes of the resonator is $c/2d$ [see Eq (6 2-21)] Hence if the laser medium is able to provide a net gain over a bandwidth $\Delta\nu$ then the laser would oscillate in a number of frequencies

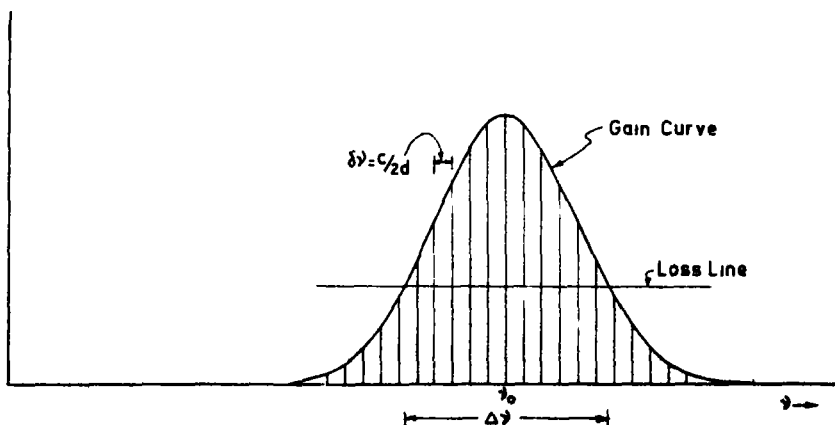



Fig 6 9 Gain profile of the active medium centered at the frequency ν_0 and of width $\Delta\nu$ The longitudinal mode separation is $\delta\nu = c/2d$

separated by $c/2d$ and the actual number of oscillating modes would be

$$N + 1 = 1 + \text{integer closest to (but less than)}\ \frac{\Delta\nu}{c/2d} \qquad (6.7\text{-}1)$$

where we have assumed that there is a mode at the center of the line The total output from the laser can be written as a superposition of the fields of all modes as

$$E(z, t) = \sum_{m=-N/2}^{N/2} A_n \exp\left[-i2\pi\nu_n\left(t - \frac{z}{c}\right) + i\phi_n\right] \qquad (6.7\text{-}2)$$

where A_n represents the amplitude corresponding to the nth mode and ϕ_n its phase In general, the amplitudes of the various modes are defined by the gain available at the various frequencies and are determined by the gain profile of the medium On the other hand, the phases ϕ_n of the various modes are arbitrary For such a case, the output laser intensity, which is proportional to $|E(z, t)|^2$ is given by

$$\begin{aligned} I(t) &= K\,|E(t)|^2 \\ &= K \sum_{n=-N/2}^{N/2} |A_n|^2 + K \sum_{n(\neq m)} \sum_m A_n A_m^* \\ &\quad \times \exp[-i2\pi(\nu_n - \nu_m)(t - z/c) + i(\phi_n - \phi_m)] \end{aligned} \qquad (6.7\text{-}3)$$

where K represents the proportionality constant For arbitrary values of the amplitudes and phases of the various modes, the above expression represents a fluctuating output intensity of the laser as plotted in Fig 6 10 It can be verified from Eq (6 7-3) that

$$I(t) = I(t + 1/\delta\nu) \qquad (6.7\text{-}4)$$

where $\delta\nu = (c/2d)$ represents the intermodal spacing Thus the intensity

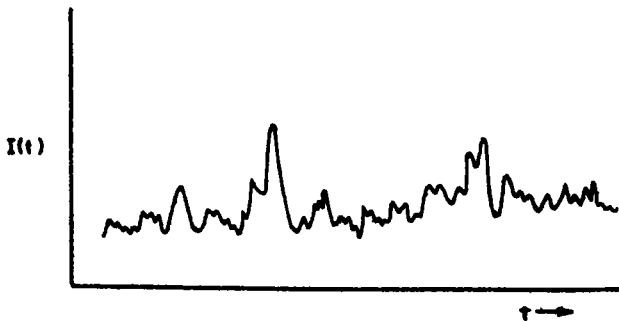



Fig. 6.10 Time variation of the output intensity of a typical multimode laser when the modes have random phases.

pattern repeats itself with a periodicity $(\delta\nu)^{-1}$ even when the modes are uncorrelated At the same time, from Eq (6 7-3) it follows that within this quasiperiodic intensity fluctuations, the shortest fluctuation occurs in a time of the order of the frequency difference between the two extreme modes and is given by

$$t_f = (\nu_0 + \tfrac{1}{2}N\,\delta\nu - \nu_0 + \tfrac{1}{2}N\,\delta\nu)^{-1} = (N\,\delta\nu)^{-1} = 1/\Delta\nu \quad (6\ 7\text{-}5)$$

i e , the inverse of the bandwidth of the laser medium When the laser is oscillating below threshold, the various modes are largely uncorrelated as a result of the absence of correlation between various spontaneously emitting sources These fluctuations become much less on passing above threshold but the different modes still remain essentially uncorrelated and the output intensity fluctuates with time From Eq (6 7-3) it follows that the average intensity is given by

$$I_{av} = K\sum_n |A_n|^2 \quad (6\ 7\text{-}6)$$

If we assume all modes to have the same amplitude (A_0) then

$$I_{av} = (N+1)KA_0^2 \quad (6\ 7\text{-}7)$$

Let us now consider the case when all the modes are locked in phase, i e , $\phi_n = \phi_0$ If we assume for simplicity that their amplitudes are also equal, then the resultant laser output amplitude will be given by

$$\begin{aligned} E(z,t) &= A_0 e^{i\phi_0} \sum_{n=-N/2}^{N/2} \exp\left[-2\pi i\nu_n\left(t-\frac{z}{c}\right)\right] \\ &= A_0 e^{i\phi_0} \exp\left[-2\pi i\nu_0\left(t-\frac{z}{c}\right)\right]\left\{\frac{\sin[\pi\,\delta\nu(N+1)(t-z/c)]}{\sin[\pi\,\delta\nu(t-z/c)]}\right\} \end{aligned} \quad (6\ 7\text{-}8)$$

Hence the output laser intensity which is proportional to $|E(z,t)|^2$ would be

$$I(t) = KA_0^2\left\{\frac{\sin[2\pi(N+1)(\delta\nu/2)(t-z/c)]}{\sin[2\pi(\delta\nu/2)(t-z/c)]}\right\}^2 \quad (6\ 7\text{-}9)$$

The variation of the intensity given by Eq (6 7-9) as a function of time at a fixed point (or as a function of z at a fixed time) is shown in Fig 6 11 As can be seen from the figure, the output is now a regular sequence of well-defined pulses From Eq (6 7-9) one may observe the following

(i) The output consists of a sequence of pulses which are separated by a time interval of $1/\delta\nu = 2d/c$ This is exactly the time taken to complete a round trip in the resonator cavity Hence, the mode-locked

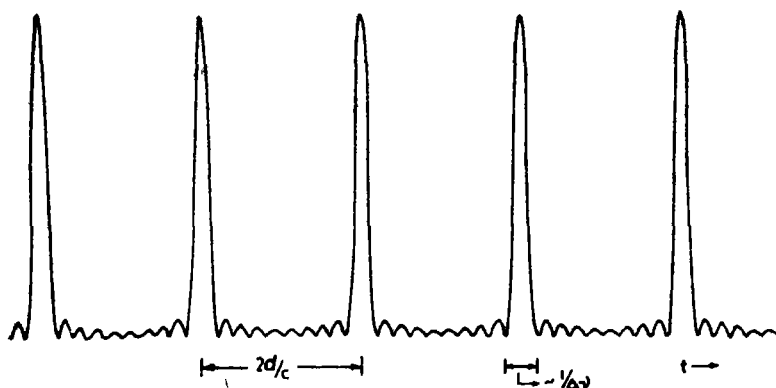



Fig 6 11 Time variation of the output intensity of a mode-locked laser The pulse repetition rate is $1/\delta\nu$ ($=2d/c$), the time taken for a complete round trip of the optical cavity and the pulse width is given by $1/\Delta\nu$, i e , the inverse of the bandwidth of the gain medium

condition can be visualized as a pulse which is traveling back and forth in the laser cavity and which loses a part of its energy through the output mirror in every round trip

(ii) From Eq (6 7-9) it also follows that the duration of the pulse is approximately given by

$$t_p \simeq \frac{1}{N\,\delta\nu} = \frac{1}{\Delta\nu} \qquad (6\,7\text{-}10)$$

i e , inverse of the bandwidth of the atomic line Thus, the larger the oscillating bandwidth of the laser medium, the smaller will be the pulse width For typical gas lasers, the pulse widths that can be obtained are about 1 nsec Since the oscillating bandwidths of solid state lasers are much larger, one can indeed obtain pulses as short as 1 psec or even smaller Such pulses are referred to as ultrashort pulses and find widespread applications in the study of ultrafast phenomena in physics, chemistry, and biology (Shapiro, 1977)

(iii) From Eq (6 7-9) one can also obtain the peak intensity of the output pulses as

$$I_{\text{peak}} = (N+1)^2 K A_0^2 \qquad (6\,7\text{-}11)$$

which is $(N+1)$ times the average intensity [see Eq (6 7-7)] Thus for typical solid state lasers having 10^3 to 10^4 modes of oscillation, the peak power enhancement obtained due to mode locking is very large

Various techniques have evolved for locking together the various modes For example, one could introduce into the resonator cavity a

device which modulates periodically either the loss or the refractive index of the cavity Such a technique is referred to as active mode locking since the modulating device is run by a source other than the laser In order to understand how, for example, a loss modulator placed inside the resonator can lock the various modes in phase, let us assume that the loss is modulated at a frequency $\delta\nu$ which is the intermodal spacing As soon as the laser is switched on, the mode that lies nearest to the line center (namely, at frequency ν_0) would start oscillating first Since the loss is modulated at the frequency $\delta\nu$, the amplitude of this mode would also oscillate at the frequency $\delta\nu$ and the resultant field will be given by

$$(A_0 + A_1 \cos 2\pi\,\delta\nu\, t)\cos 2\pi\nu_0 t$$
$$= A_0 \cos 2\pi\nu_0 t + \tfrac{1}{2}A_1 \cos 2\pi(\nu_0 + \delta\nu)t + \tfrac{1}{2}A_1 \cos 2\pi(\nu_0 - \delta\nu)t \qquad (6\ 7\text{-}12)$$

Thus this modulated central mode is a superposition of oscillating modes at frequencies ν_0, $\nu_0 + \delta\nu$, and $\nu_0 - \delta\nu$ The oscillating field at the frequencies $\nu_0 + \delta\nu$ and $\nu_0 - \delta\nu$ forces the modes corresponding to these frequencies into oscillation and thus these new modes have a perfect phase relationship with the mode at ν_0 The amplitudes of these new modes are also modulated at the frequency $\delta\nu$ and they create additional frequencies $\nu_0 - 2\,\delta\nu$ and $\nu_0 + 2\,\delta\nu$ in addition to those already present Thus all the modes are forced into oscillation in a definite phase and this leads to mode locking

Mode locking can also be obtained by using saturable absorbers

6.8. Confocal Resonator System

In this section, we shall obtain the modes of a symmetric confocal resonator system which consists of a pair of mirrors of equal radii of curvatures separated by a distance equal to the radius of curvature, see Fig 6 12 Since the resonator system is symmetric about the midplane N_1N_2, the modes of the resonator can be obtained by simply requiring that the field distribution across the plane AB (say), after completing half a round trip (i e , after traversing a distance R and getting reflected from mirror M_2) must repeat itself on the plane CD Such a condition would give us the transverse modes of the resonator The oscillation frequencies can also be obtained by requiring that the phase shift suffered by the wave in half a round trip must be equal to an integral multiple of π

In order to obtain the modes, we must first determine how a field transforms as it propagates through a distance z; this we will derive using

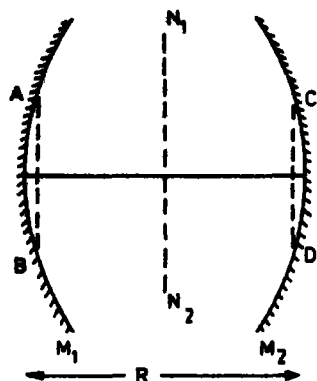



Fig 6 12 A symmetric confocal resonator system consisting of two mirrors of radii of curvatures R and separated by a distance R

Huygens' principle, for a more rigorous derivation, the reader may look up Ghatak and Thyagarajan (1978) Let $f(x, y)$ represent the field distribution on a plane, say $z = 0$ (see Fig 6 13) The field distribution on a plane at a distance z is given by a superposition of the field due to spherical waves emanating from every point on the plane $z = 0$ Let us consider an elemental area $dx'\,dy'$ centered at the point (x', y') on the plane $z = 0$ The field produced by this element at the point (x, y) on the plane z would be proportional to

$$f(x', y')\frac{e^{-ikr}}{r}\,dx'\,dy'$$

where r is the distance between the points $(x', y', 0)$ and (x, y, z)

$$r = [(x - x')^2 + (y - y')^2 + z^2]^{1/2} \qquad (6\ 8\text{-}1)$$

Thus the total field at any point (x, y) on the plane z would be given by

$$g(x, y, z) = K\iint f(x', y')\frac{e^{-ikr}}{r}\,dx'\,dy' \qquad (6\ 8\text{-}2)$$

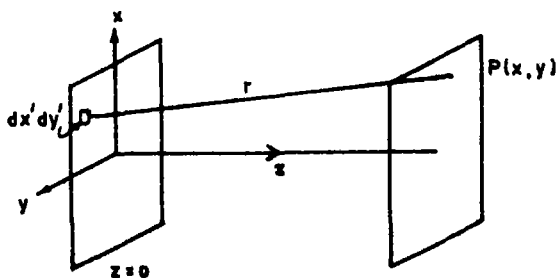



Fig. 6 13 The field distribution on the plane z is given by the superposition of the field due to spherical waves emanating from every point on the plane $z = 0$

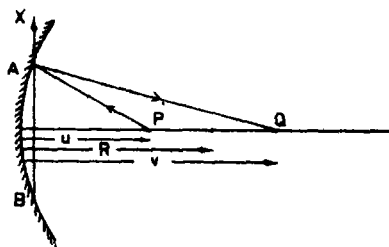



Fig 6 14 Spherical waves emanating from a point P (which is at a distance u from the mirror) after reflection from the mirror converge to the point Q which is at a distance v from the mirror

where K is a constant If we restrict ourselves to a region close to the z axis then $x, y, x', y' \ll z$ and we may write

$$r = z\left[1 + \frac{(x - x')^2 + (y - y')^2}{z^2}\right]^{1/2}$$

$$\simeq z + \frac{(x - x')^2 + (y - y')^2}{2z} \qquad (6\ 8\text{-}3)$$

Thus

$$g(x, y, z) = \frac{Ke^{-ikz}}{z} \iint f(x', y') \exp\left\{-\frac{ik}{2z}[(x - x')^2 + (y - y')^2]\right\} dx'\, dy' \qquad (6\ 8\text{-}4)$$

From a rigorous analysis one may show that $K = i/\lambda$ (see, e g , Ghatak and Thyagarajan, 1978) Given the field distribution on a plane, Eq (6 8-4) may be used to calculate the field distribution on any other plane

Before we can apply the results of the above analysis for calculating the modal patterns, we have to determine the effect on the field when it undergoes reflection from a mirror of radius R In order to calculate this, we note that a spherical wave emanating from an axial point situated at a distance u from the mirror, after getting reflected from the mirror, becomes a spherical wave converging to a point at a distance v from the mirror† (see Fig 6 14), u and v are related through the mirror equation

$$\frac{1}{u} + \frac{1}{v} = \frac{2}{R} \qquad (6\ 8\text{-}5)$$

Thus the mirror converts the incident diverging spherical wave of radius u to a converging spherical wave of radius v The phase variation produced on the plane AB by the diverging spherical wave would be given by e^{-ikr}, where $r = [x^2 + y^2 + u^2]^{1/2}$, x and y being transverse coordinates on the plane AB If we assume $x, y \ll u$ (which is the paraxial approximation),

† We are restricting ourselves to the paraxial approximation

then we may write

$$r = u\left(1 + \frac{x^2 + y^2}{u^2}\right)^{1/2} \approx u + \frac{x^2 + y^2}{2u} \tag{6 8-6}$$

Thus the phase distribution on the plane AB would be

$$\exp\left[-\frac{ik}{2u}(x^2 + y^2)\right]$$

where we have omitted the constant phase term $\exp(-iku)$ Similarly the phase distribution produced on the plane AB by the spherical wave converging to the point Q at a distance v from the mirror would be

$$\exp\left[i\frac{k}{2v}(x^2 + y^2)\right]$$

where we have again omitted the constant phase term $\exp(+ikv)$ Hence, if p_m represents the factor which when multiplied to the incident phase distribution gives the emergent phase distribution, then we have

$$p_m \exp\left[-i\frac{k}{2u}(x^2 + y^2)\right] = \exp\left[i\frac{k}{2v}(x^2 + y^2)\right]$$

or

$$p_m = \exp\left[i\frac{k}{2f}(x^2 + y^2)\right] \tag{6 8-7}$$

where $f = R/2$ represents the focal length of the mirror Equation (6 8-7) represents the effect of the mirror on the incident field distribution

A field distribution $f(x, y)$ would be a transverse mode of the resonator if it reproduces itself after traversing from plane AB to plane CD (see Fig 6 12) Thus if $f(x, y)$ represents the field distribution on the plane AB, then the field distribution on the plane CD (after half a round trip) would be given by

$$\begin{aligned} g(x, y) = \frac{i}{\lambda R} e^{-ikR} \iint_{\mathscr{A}} f(x', y') \\ \times \exp\left\{-i\frac{k}{2R}[(x - x')^2 + (y - y')^2]\right\} dx'\, dy' \\ \times \exp\left[i\frac{k}{2f}(x^2 + y^2)\right] \end{aligned} \tag{6 8-8}$$

where the integration is performed over the surface represented by

AB The field distribution $f(x, y)$ would be a mode of the resonator if

$$g(x, y) = \sigma f(x, y) \tag{6.8-9}$$

where σ is some complex constant The losses suffered by the field would be governed by the magnitude of σ, and the phase shift suffered by the wave (which determines the oscillation frequencies of the resonator) would be determined by the phase of σ Using Eq (6.8-9) Eq (6.8-8) may be written as

$$\sigma f(x, y) = \frac{\iota}{\lambda R} e^{-\iota kR} \iint_{\mathcal{A}} dx'\, dy'\, f(x', y') \times \exp\left[-\iota \frac{k}{2R}(x'^2 + y'^2 - 2xx' - 2yy')\right] \exp\left[\iota \frac{k}{2R}(x^2 + y^2)\right] \tag{6.8-10}$$

where we have used the fact that $f = R/2$

In order to solve Eq (6.8-10) we define a function $u(x, y)$ through the following relation

$$u(x, y) = f(x, y) \exp\left[-\iota \frac{k}{2R}(x^2 + y^2)\right] \tag{6.8-11}$$

We also introduce a set of dimensionless variables

$$\xi = \left(\frac{k}{R}\right)^{1/2} x = \left(\frac{2\pi}{\lambda R}\right)^{1/2} x, \qquad \eta = \left(\frac{k}{R}\right)^{1/2} y = \left(\frac{2\pi}{\lambda R}\right)^{1/2} y \tag{6.8-12}$$

Using Eqs (6.8-11) and (6.8-12), Eq (6.8-10) becomes

$$\sigma u(\xi, \eta) = \frac{1}{2\pi} e^{-\iota kR} \iint_{\mathcal{B}} d\xi'\, d\eta'\, u(\xi', \eta') e^{\iota(\xi\xi' + \eta\eta')} \tag{6.8-13}$$

where $\mathcal{B}$ represents the modified limits of integration

In order to simplify the analysis, we assume the mirrors to be rectangular with dimensions $2a \times 2b$ In such a case we may write

$$\sigma u(\xi, \eta) = \frac{\iota}{2\pi} e^{-\iota kR} \int_{-\xi_0}^{+\xi_0} \int_{-\eta_0}^{+\eta_0} u(\xi', \eta') e^{\iota(\xi\xi' + \eta\eta')}\, d\xi'\, d\eta' \tag{6.8-14}$$

where $\xi_0 = (k/R)^{1/2}a$ $\eta_0 = (k/R)^{1/2}b$ In order to solve Eq (6.8-14), we

try the separation of variables technique and write

$$\sigma = \kappa\tau \qquad u(\xi, \eta) = p(\xi)q(\eta) \tag{6.8-15}$$

On substituting Eq. (6.8-15) in Eq. (6.8-14), we find that the variables indeed separate out and we obtain

$$\kappa p(\xi) = \left(\frac{i}{2\pi}\right)^{1/2} e^{-ikR/2} \int_{-\xi_0}^{+\xi_0} p(\xi')e^{i\xi\xi'}\, d\xi' \tag{6.8-16}$$

$$\tau q(\eta) = \left(\frac{i}{2\pi}\right)^{1/2} e^{-ikR/2} \int_{-\eta_0}^{+\eta_0} q(\eta')e^{i\eta\eta'}\, d\eta' \tag{6.8-17}$$

The integrals appearing in Eqs. (6.8-16) and (6.8-17) are referred to as finite Fourier transforms; they reduce to the usual Fourier transforms in the limit $\xi_0 \to \infty$ and $\eta_0 \to \infty$. It has been shown by Slepian and Pollack (1961) that the solutions of Eqs. (6.8-16) and (6.8-17) are prolate spheroidal functions. We will only consider the case when $\xi_0 = (2\pi)^{1/2}(a^2/\lambda R)^{1/2} \gg 1$ and $\eta_0 = (2\pi)^{1/2}(b^2/\lambda R)^{1/2} \gg 1$, i.e., resonators having large Fresnel numbers† N_1 ($= a^2/\lambda R$) and N_2 ($= b^2/\lambda R$). For such a case we may extend the limits of integration in Eqs. (6.8-16) and (6.8-17) from $-\infty$ to ∞. Thus Eq. (6.8-16) becomes

$$Ap(\xi) = \int_{-\infty}^{+\infty} p(\xi')e^{i\xi\xi'}\, d\xi' \tag{6.8-18}$$

where

$$A = \kappa\left(\frac{2\pi}{i}\right)^{1/2} e^{ikR/2} \tag{6.8-19}$$

An identical equation is satisfied by $q(\eta)$. Equation (6.8-18) requires that (apart from some constant factors) $p(\xi)$ be its own Fourier transform.

In order to solve Eq. (6.8-18) for $p(\xi)$ we differentiate Eq. (6.8-18) twice with respect to ξ and obtain

$$A\frac{d^2p}{d\xi^2} = -\int_{-\infty}^{+\infty} p(\xi')\xi'^2 e^{i\xi\xi'}\, d\xi' \tag{6.8-20}$$

† When the field gets reflected at the mirror, it is restricted in the tranverse dimension and hence it undergoes diffraction divergence as it propagates back to the first mirror. The angle of divergence ~ λ/a. The angle subtended by one of the mirrors at the other mirror is ~ a/R, R being the distance between the mirrors; thus for the diffraction losses to be low $a/R \gg \lambda/a$ or $a^2/\lambda R \gg 1$.

We now consider the integral

$$I = \int_{-\infty}^{+\infty} \frac{d^2p}{d\xi'^2} e^{i\xi\xi'}\, d\xi' \tag{6.8-21}$$

For the mode, we assume $p(\xi)$ and its derivative to vanish at infinity and integrate Eq (6.8-21) twice by parts and obtain

$$\begin{aligned} \int_{-\infty}^{+\infty} \frac{d^2p}{d\xi'^2} e^{i\xi\xi'}\, d\xi' &= -\xi^2 \int_{-\infty}^{+\infty} p(\xi')e^{i\xi\xi'}\, d\xi' \\ &= -A\xi^2 p(\xi) \end{aligned} \tag{6.8-22}$$

Combining Eqs (6.8-20) and (6.8-22), we get

$$A\left[\frac{d^2p}{d\xi^2} - \xi^2 p(\xi)\right] = \int_{-\infty}^{+\infty} \left(\frac{d^2p}{d\xi'^2} - \xi'^2 p\right)e^{i\xi\xi'}\, d\xi' \tag{6.8-23}$$

Comparing Eq (6.8-23) with Eq (6.8-18), we note that both $p(\xi)$ and $[d^2p/d\xi^2 - \xi^2 p(\xi)]$ satisfy the same equation and as such one must have

$$\frac{d^2p}{d\xi^2} - \xi^2 p(\xi) = -Kp(\xi) \tag{6.8-24}$$

where K is some constant We may rewrite Eq (6.8-24) as

$$\frac{d^2p}{d\xi^2} + [K - \xi^2]p(\xi) = 0 \tag{6.8-25}$$

The solutions of Eq (6.8-25) with the condition that $p(\xi)$ vanish at large values of ξ are the Hermite–Gauss functions†

$$p_m(\xi) = N_m H_m(\xi) e^{-\xi^2/2} \tag{6.8-26}$$

where N_m is the normalization constant $H_m(\xi)$ represents the mth-order Hermite polynomial of argument ξ A few lower-order Hermite polynomials are

$$H_0(\xi) = 1, \qquad H_1(\xi) = 2\xi, \qquad H_2(\xi) = 4\xi^2 - 2, \tag{6.8-27}$$

Thus the complete solution of Eq (6.8-10) may be written as

$$f(x, y) = CH_m(\xi)H_n(\eta)e^{-(\xi^2+\eta^2)/2}e^{+i(\xi^2+\eta^2)/2} \tag{6.8-28}$$

where C is some constant, here m and n represent the transverse mode numbers and determine the transverse field distribution of the mode The

† See Section 2.8

Hermite–Gauss functions satisfy the equation

$$i^m H_m(\xi)e^{-\xi^2/2} = (2\pi)^{-1/2}\int_{-\infty}^{+\infty} d\xi'\, H_m(\xi')e^{-\xi'^2/2}e^{i\xi\xi'} \tag{6.8-29}$$

If we use the fact that $p(\xi)$ in Eq. (6.8-16) is a Hermite–Gauss function, then using Eq. (6.8-29), one readily obtains

$$\kappa = i^m(i)^{1/2}e^{-ikR/2} = \exp\left\{-i\left[\frac{kR}{2} - (m + \tfrac{1}{2})\frac{\pi}{2}\right]\right\} \tag{6.8-30}$$

Similarly

$$\tau = i^n(i)^{1/2}e^{-ikR/2} = \exp\left\{-i\left[\frac{kR}{2} - (n + \tfrac{1}{2})\frac{\pi}{2}\right]\right\} \tag{6.8-31}$$

Thus from Eq. (6.8-15) we have

$$\sigma = \kappa\tau = \exp\left\{-i\left[kR - (m + n + 1)\frac{\pi}{2}\right]\right\} \tag{6.8-32}$$

Observe that $|\sigma| = 1$ implying the absence of any losses. This can be attributed to the fact that, in our analysis, we have essentially assumed the mirrors to be of extremely large transverse dimensions.

Since the phase of σ represents the phase shift suffered by the wave in half a round trip, one must have

$$kR - (m + n + 1)\pi/2 = q\pi, \qquad q = 1, 2, 3, \ldots \tag{6.8-33}$$

where q refers to the longitudinal mode number. Using $k = 2\pi\nu/c$ we obtain the frequencies of oscillation of the cavity as

$$\nu_{mnq} = (2q + m + n + 1)\frac{c}{4R} \tag{6.8-34}$$

Observe that all modes having the same value of $(2q + m + n)$ would have the same oscillation frequency and hence would be degenerate. The frequency separation between two modes having the same value of m and n but adjacent values of q is [see also Eq. (6.2-21)],

$$\Delta\nu_q = \frac{c}{2R} \tag{6.8-35}$$

The frequency separation between two transverse modes corresponding to the same value of q is

$$\Delta\nu_m \text{ (or } \Delta\nu_n) = \frac{c}{4R} \tag{6.8-36}$$

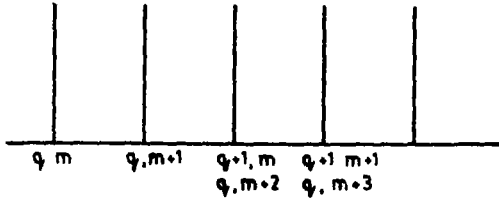



Fig 6 15 The longitudinal mode spectrum of a confocal resonator The frequency separation between two transverse modes is half that between two longitudinal modes

which is half that between two consecutive longitudinal modes [cf Eq (6 2-22)]—see Fig 6 15

In our analysis $|\sigma| = 1$, i e , there are no diffraction losses Boyd and Gordon (1961) have used a more rigorous analysis and have shown that the diffraction losses of the confocal geometry are much smaller than for the plane-parallel resonator

In Fig 6 16 we depict the transverse intensity distribution corresponding to the mode amplitude distributions given by Eq (6 8-26) Figure 6 4 shows the photographs of the intensity distribution corresponding to different transverse modes of a laser resonator Observe that higher-order modes extend more in the transverse dimension and hence would have higher diffraction losses

The field given by Eq (6 8-28) is the field distribution in a transverse plane passing through the pole of the mirror The field distribution at any other plane can be derived by using the diffraction formula given by Eq

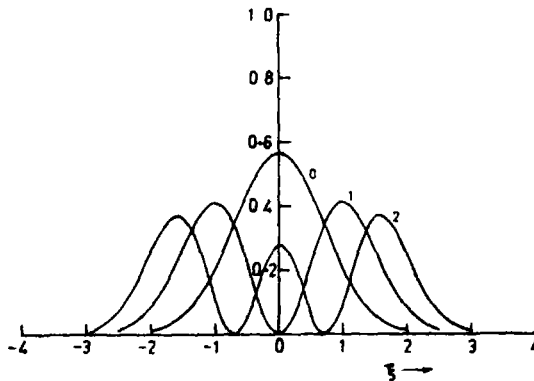



Fig 6 16 Transverse intensity distributions corresponding to the three lowest-order Hermite–Gauss modes Observe that higher-order modes have a more spread out intensity pattern The curves labeled 0, 1, and 2 correspond to $p_0^2(\xi)$ $p_1^2(\xi)$, and $p_2^2(\xi)$ respectively [see Eq. (6.8-26)]

(6 8-4) Thus the field distribution midway between the two mirrors is

$$f_M(x, y) = \frac{iC}{\pi} e^{-ikR/2} \iint H_m(\xi')H_n(\eta') \exp[-\tfrac{1}{2}\xi'^2(1 - i) - \tfrac{1}{2}\eta'^2(1 - i)]$$
$$\times \exp[-i(\xi^2 + \xi'^2 - 2\xi\xi') - i(\eta^2 + \eta'^2 - 2\eta\eta')] \, d\xi' \, d\eta'$$
$$= \frac{iC}{\pi} e^{-ikR/2} e^{i(\xi^2+\eta^2)} \int H_m(\xi') \exp[-\tfrac{1}{2}\xi'^2(1 + i) + 2i\xi\xi'] \, d\xi'$$
$$\times \int H_n(\eta') \exp[-\tfrac{1}{2}\eta'^2(1 + i) + 2i\eta\eta'] \, d\eta' \qquad (6\ 8\text{-}37)$$

Using the standard integral

$$\int_{-\infty}^{+\infty} e^{-x^2+2xy} H_n(\alpha x) \, dx = \pi^{1/2} e^{y^2} (1 - \alpha^2)^{n/2} H_n\left[\frac{\alpha y}{(1 - \alpha^2)^{n/2}}\right] \qquad (6\ 8\text{-}38)$$

we obtain

$$f_M(x, y) = C \exp\left[-i\left(\frac{kR}{2} - \frac{\pi}{2}\right)\right] \exp[-(\xi^2 + \eta^2)] H_m(2^{1/2}\xi) H_n(2^{1/2}\eta) \qquad (6\ 8\text{-}39)$$

which shows that the phase along the transverse plane midway between the mirrors is constant Thus the phase fronts are plane midway between the mirrors From Eq (6 8-28) it can also be seen that the phase front of the modal field distribution has a radius of curvature R which is equal to the radius of the resonator mirror

6 9 Planar Resonators

A rigorous analysis of the modes of a plane parallel resonator was first performed by Fox and Li (1961), using the scalar approximation In such an approximation they assumed that the field is nearly transverse and polarized along one direction Such an analysis is valid as long as the mirror sizes are much larger than the wavelength of light The analysis by Fox and Li (1961) consisted of assuming a certain field configuration at one of the mirrors of the resonator and then a calculation of the field produced at the second mirror The field at the second mirror is used to calculate back the field distribution on the first mirror Such a computation is repeated over a large number of traversals Fox and Li (1961) showed that after many traversals, the relative field distribution settles down to a steady pattern, i e , it does not vary from transit to transit but

only the amplitude of the field decays exponentially Such a field distribution corresponds to the normal mode of the resonator By changing the initial condition on the first mirror, other modes can also be obtained

It was shown in Section 6 8 that if $U(x_1, y_1)$ is the field distribution on a plane $z = 0$, then the field distribution at a point (x, y, z) would be given by Eq (6 8-4) Thus is the distance between the mirrors is d, then the field on the second mirror can easily be obtained from Eq (6 8-4) by replacing z by d Using the resulting field distribution, Eq (6 8-4) can again be used for calculating the field distribution on the first mirror

If the field distribution $U(x, y)$ corresponds to the mode of the resonator, it must be the same on both the mirrors (since they are identical) Thus the field on the plane $z = d$ must be proportional to $U(x, y, 0)$ where the constant of proportionality σ could, in general, be complex Thus, for a mode, we must have

$$\sigma U(x, y) = \frac{i}{\lambda d} e^{-ikd} \iint U(x_1, y_1) \times \exp\left\{-\frac{i\pi}{\lambda d}[(x - x_1)^2 + (y - y_1)^2]\right\} dx_1\, dy_1 \quad (6\ 9\text{-}1)$$

As discussed in Section 6 8, for mirrors of rectangular geometry Eq (6 9-1) can be written as two equations by using the separation of variables technique We write

$$\sigma = \kappa\tau \quad (6\ 9\text{-}2)$$

and try a separable solution of the form

$$U(x, y) = f(x)g(y) \quad (6\ 9\text{-}3)$$

If we substitute Eqs (6 9-2) and (6 9-3) in Eq (6 9-1), we find that the variables indeed separate out and we obtain

$$\kappa f(x) = e^{-ikd/2} i^{1/2} \int_{-(N^{1/2})}^{N^{1/2}} f(\xi_1) \exp[-i\pi(\xi - \xi_1)^2]\, d\xi_1 \quad (6\ 9\text{-}4)$$

and

$$\tau g(y) = e^{-ikd/2} i^{1/2} \int_{-(N^{1/2})}^{N^{1/2}} g(\eta_1) \exp[-i\pi(\eta - \eta_1)^2]\, d\eta_1 \quad (6\ 9\text{-}5)$$

where

$$\xi = \frac{x}{(\lambda d)^{1/2}} = N^{1/2}\frac{x}{a} \quad (6\ 9\text{-}6)$$

$$\eta = \frac{y}{(\lambda d)^{1/2}} = N^{1/2}\frac{y}{a} \quad (6\ 9\text{-}7)$$

and $N = a^2/\lambda d$ is the Fresnel number Equations (6 9-4) and (6 9-5) were solved by Fox and Li (1961) with the help of a computer They showed that if the initial field distribution on one mirror is assumed to be uniform and the phase is constant across the mirror, after about 300 transists of the field through the resonator, the relative field distribution

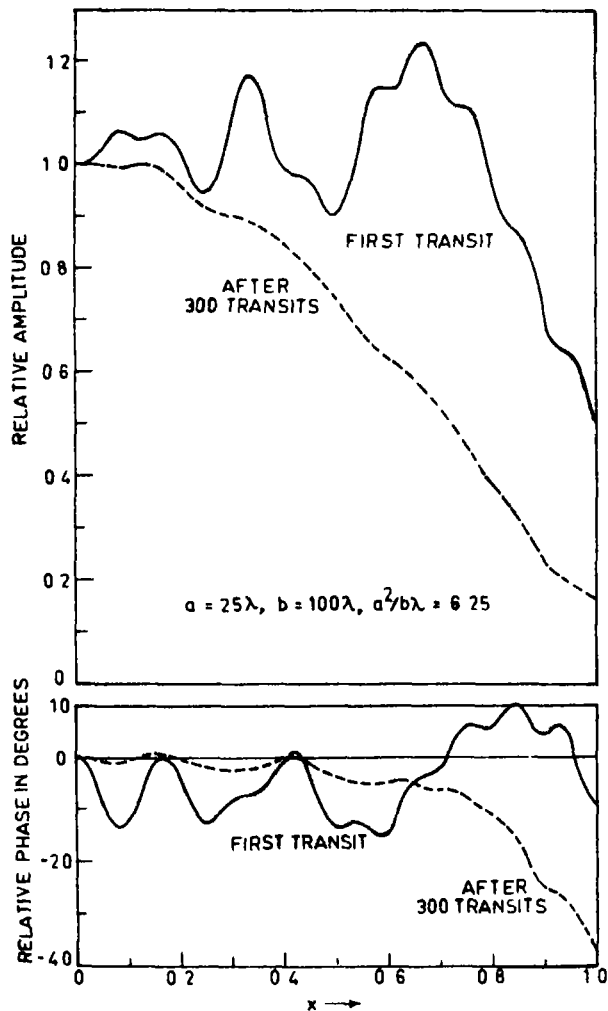



Fig 6 17 Relative amplitude and phase distributions of field intensity distributions at the mirrors (for infinite strip mirrors) after the first transit and 300 transits The amplitude and phase distributions stabilize after 300 transits Here a represents the halfwidth of the mirrors and b the separation between the mirrors (After Fox and Li 1961)

settles down to a steady value whose amplitude is a maximum at the center and decreases monotonically as one moves away from the center The relative amplitude distribution across the mirror after a single transit and after 300 transits for a resonator consisting of a pair of infinite strip mirrors is shown in Fig 6 17 Also shown is the relative phase distribution across the mirror after one and 300 transits By observing the amplitude and phase at any point on the wave front (after attainment of steady state), Fox and Li could determine the power loss and the phase shift per transit of the mode Similarly Fox and Li (1961) could obtain the lowest-order antisymmetric mode also, this mode was obtained by launching a wave for which the field intensity over the mirror is constant but there is a phase difference of π between the two halves of the mirror

6 10 General Spherical Resonator

In Section 6 8 we had discussed in detail the symmetric confocal resonator system In general, a resonator could consist of a pair of spherical mirrors of radii of curvatures R_1 and R_2 which are separated by a distance d If we can find a Guassian beam which has a radius of curvature R_1 at the first mirror R_2 at the second mirror and whose beam size at both the mirrors is much smaller than the mirror size then such a Gaussian beam can resonate in the resonator this is shown in Fig 6 18

In order to calculate the diffraction of a Gaussian beam in space we consider a Gaussian beam on the plane $z = 0$ given by

$$u(x\ y\ 0) = A \exp\left(-\frac{x^2 + y^2}{w_0^2}\right) \tag{6 10-1}$$

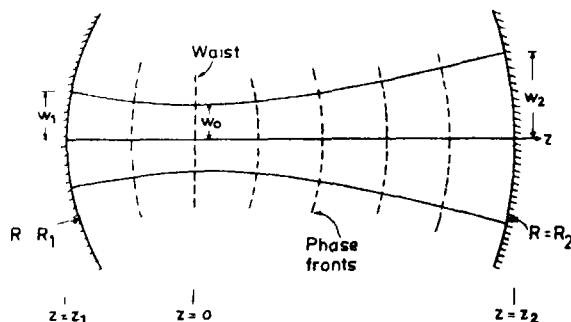



Fig 6 18 A resonator formed of two spherical mirrors of radii R_1 and R_2 and separated by a distance d A Gaussian beam with phase front radii R_1 and R_2 at the mirrors M_1 and M_2 and whose width is much smaller than the mirror sizes can resonate in the resonator

Using Eq (6 8-4) one can calculate the amplitude distribution at any plane z as

$$u(x, y, z) = A\frac{i\pi}{\lambda}\frac{2w_0^2}{2z + ikw_0^2} \times \exp\left\{-ik\left[z + \frac{x^2+y^2}{R(z)}\right]\right\}\exp\left[-\frac{x^2+y^2}{w^2(z)}\right] \quad (6\ 10\text{-}2)$$

where

$$R(z) = z\left(1 + \frac{\pi^2 w_0^4}{\lambda^2 z^2}\right) \quad (6\ 10\text{-}3)$$

and

$$w(z) = w_0\left(1 + \frac{\lambda^2 z^2}{\pi^2 w_0^4}\right)^{1/2} \quad (6\ 10\text{-}4)$$

represent, respectively the radius of curvature and the beam size of the Gaussian beam at the plane z As is evident, a Gaussian beam remains Gaussian as it propagates and the beam which had a plane phase front at the plane $z = 0$ [see Eq (6 10-1)] becomes spherical with a radius $R(z)$ given by Eq (6 10-3) The plane $z = 0$ (where the phase front is plane and the beam has the minimum width) is referred to as the waist of the Gaussian beam The beam size increases on either side of the waist as given by Eq (6 10-4)

In order to determine the parameters of the Gaussian beam that would resonate inside a general spherical resonator, we assume that the poles of the mirrors with radii of curvatures R_1 and R_2 are at $z = z_1$ and $z = z_2$ respectively (see Fig 6 18) We assume that a Gaussian beam of waist size w_0 resonates between the mirrors Using Eq (6 10-3) the radius of curvature of the phase front of the Gaussian beam at $z = z_1$ and $z = z_2$ is

$$R(z_1) = z_1 + \frac{\alpha^2}{z_1}, \qquad R(z_2) = z_2 + \frac{\alpha^2}{z_2} \quad (6\ 10\text{-}5)$$

where

$$\alpha^2 = \tfrac{1}{4}k^2 w_0^4 = \frac{\pi^2 w_0^4}{\lambda^2} \quad (6\ 10\text{-}6)$$

If $R(z_1) = R_1$ and $R(z_2) = R_2$, i e , the radius of the phase front is equal to the radii of the mirrors, then the Gaussian beam is incident normally on the mirrors, and in such a case, the beam retraces its path and the beam resonates inside the resonator We choose a sign convention that

the radius of curvature is positive if the mirror is concave towards the resonator Under such a convention $R(z_2) = R_2$ and $R(z_1) = -R_1$ Thus

$$z_1 + \frac{\alpha^2}{z_1} = -R_1, \qquad z_2 + \frac{\alpha^2}{z_2} = R_2 \tag{6 10-7}$$

We also have

$$z_2 - z_1 = d \tag{6 10-8}$$

We now solve Eqs (6 10-7) and (6 10-8) for z_1, z_2, and α From Eq (6 10-7) we have

$$\alpha^2 = -(R_1 + z_1)z_1 = (R_2 - z_2)z_2 \tag{6 10-9}$$

Using Eq (6 10-8) we obtain

$$z_2 = \frac{d(d - R_1)}{2d - R_1 - R_2} \tag{6 10-10}$$

We now introduce two parameters

$$g_1 = 1 - \frac{d}{R_1}, \qquad g_2 = 1 - \frac{d}{R_2} \tag{6 10-11}$$

and Eq (6 10-10) becomes

$$z_2 = \frac{d(1 - g_2)g_1}{g_1 + g_2 - 2g_1g_2} \tag{6 10-12}$$

Similarly we obtain

$$z_1 = -\frac{d(1 - g_1)g_2}{g_1 + g_2 - 2g_1g_2} \tag{6 10-13}$$

$$\alpha^2 = \frac{d^2(1 - g_1g_2)g_1g_2}{(g_1 + g_2 - 2g_1g_2)^2} \tag{6 10-14}$$

Thus given a resonator with R_1, R_2, and d one can determine the parameters of the Gaussian beam that can resonate in the resonator

Knowing α^2 from Eq (6 10-14), one can calculate w_0^2 from Eq (6 10-6) and subsequently one can evaluate the beam widths at the two mirrors from Eq (6 10-4) and the values of z_1 and z_2 Thus from Eqs (6 10-6) and (6 10-14), we obtain

$$w_0^4 = \frac{\lambda^2}{\pi^2}\frac{d^2(1 - g_1g_2)g_1g_2}{(g_1 + g_2 - 2g_1g_2)^2} \tag{6 10-15}$$

Hence using Eqs (6 10-4) and (6 10-13) we get

$$w^2(z_1) = \frac{\lambda d}{\pi}\left[\frac{g_2}{g_1(1 - g_1 g_2)}\right]^{1/2} \quad (6\ 10\text{-}16)$$

Similarly, we obtain

$$w^2(z_2) = \frac{\lambda d}{\pi}\left[\frac{g_1}{g_2(1 - g_1 g_2)}\right]^{1/2} \quad (6\ 10\text{-}17)$$

$w^2(z_1)$ and $w^2(z_2)$ give the beam widths of the Gaussian mode at the two mirrors Since most of the energy in a Gaussian beam is contained within a radius of about twice the beam width, if the transverse dimensions of the mirrors are large compared to $w(z_1)$ and $w(z_2)$ then most of the energy is reflected back and the loss due to diffraction spill-over from the edges of the mirrors is small It follows from Eqs (6 10-16) and (6 10-17) that as $g_1 g_2 \to 0$ or $g_1 g_2 \to 1$, then the beamwidth at either or both the mirrors becomes infinite and our analysis breaks down. Also, for the existence of a stable Gaussian mode

$$0 \leq g_1 g_2 = \left(1 - \frac{d}{R_1}\right)\left(1 - \frac{d}{R_2}\right) \leq 1 \quad (6\ 10\text{-}18)$$

The same condition will be derived in Section 6 11 for the resonator to be stable

6 11 Geometrical Optics Analysis of Optical Resonators

In this section we perform a geometrical optics analysis of a general spherical resonator bound by two mirrors of radii of curvatures R_1 and R_2 and separated by a distance d For resonators having large Fresnel numbers (~ 100) the loss occurring due to diffraction is small (see footnote on p 134) and one can study some aspects of resonators using geometrical optics We will use the matrix treatment of geometrical optics as it is particularly suited to the study of ray propagation in an optical resonator We will obtain the condition to be satisfied by R_1, R_2, and d so that the resonator is stable In a stable resonator, a ray of light may keep bouncing between the two mirrors indefinitely because of the periodic focusing provided by the spherical mirrors On the contrary, in an unstable resonator, the ray escapes from the resonator after a few to and fro passes Before we consider the resonator, let us obtain the matrices which represent propagation of a ray through a homogeneous medium and the reflection of a ray from a concave mirror

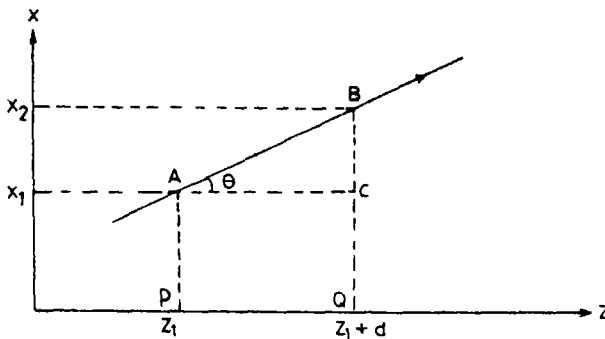



Fig 6 19 A ray propagates along a straight line in a homogeneous medium The coordinates of the ray at the point B and the coordinates of the ray at the point A are related through the translation matrix T given by Eq (6 11-4)

Let us consider a ray propagating in a homogeneous medium in the x–z plane as shown in Fig 6 19 A ray may be described by specifying the height of the ray (i e , the x coordinate) and the angle it makes with the z axis, namely, θ The z axis is usually chosen along the axis of symmetry of the optical system Thus, for a simple lens, it will be along the axis of the lens In our analysis we will be restricting ourselves to only those rays which are traveling near the z axis and which do not make a large angle with it This approximation is termed the paraxial approximation Under this approximation, both x and θ are small and we may approximate $\sin\theta \simeq \tan\theta \simeq \theta$

Let x_1 and x_1' $(= dx/dz)$ be the position and slope of the ray at a plane $z = z_1$ We wish to find the position and slope of the ray at another plane $z = z_1 + d$ lying in the same medium From Fig 6 19 it follows that

$$x_2' = \left.\frac{dx}{dz}\right|_{z_1+d} = \left.\frac{dx}{dz}\right|_{z_1} = x_1' \qquad (6\ 11\text{-}1)$$

and

$$\begin{aligned} x_2 &= BQ = BC + CQ = AC\tan\theta + AP \\ &= x_1 + x_1'd \end{aligned} \qquad (6\ 11\text{-}2)$$

where $x_1' = \tan\theta$ In deriving the first of the above relations we have used the fact that rays travel along a straight line in a homogeneous medium Thus, if we represent the initial ray by the column matrix

$$\begin{pmatrix} x_1 \\ x_1' \end{pmatrix}$$

and the final ray by the column matrix

$$\begin{pmatrix} x_2 \\ x'_2 \end{pmatrix}$$

then it follows from Eqs (6 11-1) and (6 11-2) and the rule of matrix multiplication that

$$\begin{pmatrix} x_2 \\ x'_2 \end{pmatrix} = \begin{pmatrix} 1 & d \\ 0 & 1 \end{pmatrix}\begin{pmatrix} x_1 \\ x'_1 \end{pmatrix} \tag{6 11-3}$$

Thus, the coordinates of a ray at a plane can be found from the coordinates of the ray in a previous plane at a distance d by multiplying by the matrix

$$T = \begin{pmatrix} 1 & d \\ 0 & 1 \end{pmatrix} \tag{6 11-4}$$

The above matrix is called the translation matrix and represents the effect of propagation through a homogeneous medium of length d Notice that

$$\det T = \begin{vmatrix} 1 & d \\ 0 & 1 \end{vmatrix} = 1 \tag{6 11-5}$$

where det T stands for the determinant of the matrix T

Let us now obtain the matrix which would correspond to the reflection of a ray from a concave mirror Let the coordinates of a ray be x_1, x'_1 before suffering reflection from a mirror of radius of curvature R (see Fig 6 20) We have to find the coordinates of the ray just after reflection It is obvious that the reflection from the mirror only changes the slope of the ray its position coordinate remains the same Hence

$$x_2 = x_1 \tag{6 11-6}$$

In order to determine the slope of the ray after reflection we make use of the mirror formula

$$\frac{1}{u} + \frac{1}{v} = \frac{2}{R} \tag{6 11-7}$$

where u and v represent the object and image distances from the mirror respectively (see Fig 6 20) A positive value of R corresponds to a concave mirror and a negative value of R to a convex mirror We always measure the slope of the ray with respect to the forward direction of travel Thus the slope of the initial ray would be $x'_1 = x_1/u$ (see Fig 6 20), where x_1 is the height of the ray on the mirror Similarly the slope of the reflected ray would be x'_2 $(= -x_1/v)$ Thus from Eq (6 11-7), we

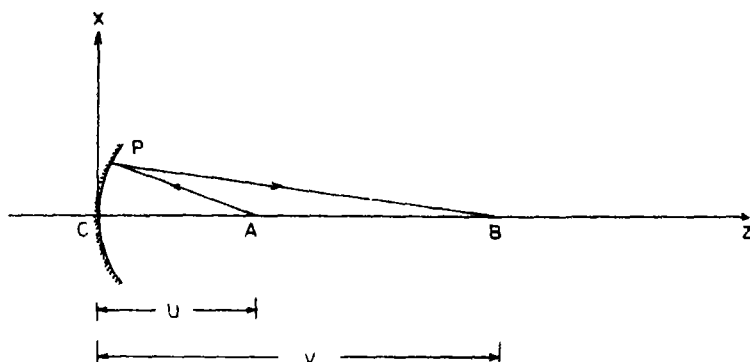



Fig 6 20 Reflection of a ray from a concave mirror The effect of such a reflection is specified by the reflection matrix $\mathcal{R}$ given by Eq (6 11 9)

obtain

$$x_1' - x_2' = \frac{2x_1}{R}$$

or

$$x_2' = x_1' - \frac{2x_1}{R} \tag{6 11-8}$$

Hence from Eqs (6 11-6) and (6 11-8), we may write the matrix for reflection from a concave mirror as†

$$\mathcal{R} = \begin{pmatrix} 1 & 0 \\ -2/R & 1 \end{pmatrix} \tag{6 11-9}$$

Observe that $\det \mathcal{R} = 1$

Having obtained the matrices for translation and reflection from a mirror, we can obtain the ray coordinates after any number of reflections and traversals by simply multiplying the matrices together In this operation care must be exercised in multiplying the matrices in the correct order Thus, let the coordinates of a ray be x_1 and x_1' in a plane which is

† In a manner similar to that employed here one may obtain the matrix for refraction at a surface separating two media or the matrix for refraction through a lens Combinations of lenses may also be treated in a similar manner Once the matrices for translation and refraction through a curved surface have been obtained, the ray can be easily traced through any complex optical system The formulation of optics using matrices is quite powerful and is well suited for computer calculations also For a more detailed analysis of matrix method in optics see e g Nussbaum (1968) Gerrard and Burch (1975) Brouwer (1964) etc

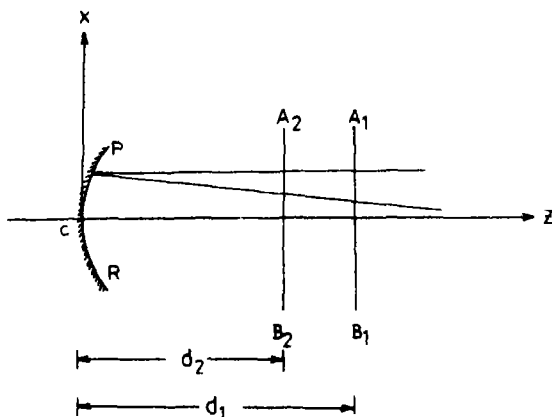



Fig 6 21 The matrix connecting the coordinates of the ray in the plane A_1B_1 and A_2B_2 is given by Eq (6 11-13)

at a distance d_1 from a concave mirror of radius R (see Fig 6 21) Let the ray suffer a reflection at the mirror and we wish to determine the coordinates of the ray in a plane at a distance d_2 from the mirror The coordinates of the ray on the mirror would be given by

$$\begin{pmatrix} \xi_1 \\ \xi_1' \end{pmatrix} = \begin{pmatrix} 1 & d_1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_1' \end{pmatrix} \tag{6 11-10}$$

where x_1 and x_1' represent the coordinates of the ray in the plane A_1B_1 The coordinates of the ray after reflection would be

$$\begin{pmatrix} \xi_2 \\ \xi_2' \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -2/R & 1 \end{pmatrix} \begin{pmatrix} \xi_1 \\ \xi_1' \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -2/R & 1 \end{pmatrix} \begin{pmatrix} 1 & d_1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_1' \end{pmatrix} \tag{6 11-11}$$

Finally, the coordinates of the ray on the plane at a distance d_2 from the mirror would hence be

$$\begin{pmatrix} x_2 \\ x_2' \end{pmatrix} = \begin{pmatrix} 1 & d_2 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \xi_2 \\ \xi_2' \end{pmatrix} = \begin{pmatrix} 1 & d_2 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -2/R & 1 \end{pmatrix} \begin{pmatrix} 1 & d_1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_1' \end{pmatrix} \tag{6 11-12}$$

Hence the matrix for the traversal from the plane A_1B_1 to A_2B_2 via the mirror would be

$$S = \begin{pmatrix} 1 & d_2 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -2/R & 1 \end{pmatrix} \begin{pmatrix} 1 & d_1 \\ 0 & 1 \end{pmatrix} \tag{6 11-13}$$

Notice that the matrices for various operations appear in reverse order S represents the system matrix between the planes A_1B_1 and A_2B_2 The system matrix is, in general, written as

$$S = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \qquad (6\,11\text{-}14)$$

Since the determinant of the product of matrices is the product of their determinants and since det $T = 1$ and det $\mathfrak{R} = 1$, we obtain the condition that for the system matrix S,

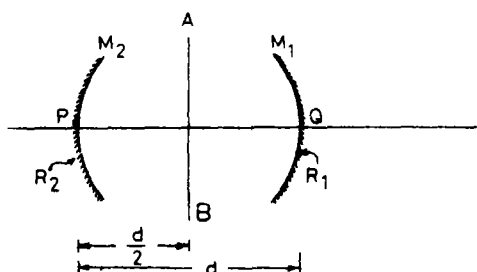
$$\det S = AD - BC = 1 \qquad (6\,11\text{-}15)$$

We will now consider a resonator bounded by two concave mirrors of radii of curvatures R_1 and R_2 and separated by a distance d (see Fig 6 22) A ray which starts from any plane AB (assumed here to be midway between the mirrors) towards mirror M_1 after reflection from it travels towards M_2 After suffering a second reflection at M_2 it returns to the plane AB This is referred to as one complete cycle of oscillation of the ray This process repeats itself and one obtains rays traveling along both directions in between the mirrors M_1 and M_2 A resonator would be stable if even after an arbitrary number of traversals the ray remains close to the axis When we impose this condition on the resonator system, we would obtain the stability condition

Let

$$S = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \qquad (6\,11\text{-}16)$$

represent the matrix for one complete traversal of a ray starting from a plane AB Thus, if the coordinates of a ray are x_0 and x_0' when it started from the plane AB, after one traversal the coordinates of the ray would

Fig 6 22 A resonator made up of mirrors of radii of curvatures R_1 and R_2 and separated by a distance d We consider an arbitrary plane AB from which a ray starts—here for simplicity the plane AB is chosen midway between the two mirrors The resonator would be stable if even after an arbitrarily large number of traversals the ray remains confined within the resonator

be given by

$$\begin{pmatrix} x_1 \\ x_1' \end{pmatrix} = \begin{pmatrix} A & B \\ C & D \end{pmatrix}\begin{pmatrix} x_0 \\ x_0' \end{pmatrix}$$

Similarly, the coordinates of the ray after two traversals would be

$$\begin{pmatrix} x_2 \\ x_2' \end{pmatrix} = \begin{pmatrix} A & B \\ C & D \end{pmatrix}\begin{pmatrix} x_1 \\ x_1' \end{pmatrix} = \begin{pmatrix} A & B \\ C & D \end{pmatrix}^2\begin{pmatrix} x_0 \\ x_0' \end{pmatrix}$$

Hence, the coordinates of the ray after n complete traversals would be

$$\begin{pmatrix} x_n \\ x_n' \end{pmatrix} = \begin{pmatrix} A & B \\ C & D \end{pmatrix}^n\begin{pmatrix} x_0 \\ x_0' \end{pmatrix} \tag{6 11-17}$$

For the resonator to be stable, we require that for a ray starting with coordinates (x_0, x_0'), the coordinates after n traversals namely, (x_n, x_n') should not diverge as n increases In order to obtain the stability criterion, we have to look at the nth power of the system matrix S for which we diagonalize the matrix

A matrix U is said to diagonalize a matrix S if

$$U^{-1}SU = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \tag{6 11-18}$$

where U^{-1} is the inverse of the matrix U defined by

$$U^{-1}U = UU^{-1} = I \tag{6 11-19}$$

I being the identity matrix

$$I \equiv \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{6 11-20}$$

λ_1 and λ_2 are called the eigenvalues of the matrix S In Eq (6 11-18) if we premultiply by U and postmultiply by U^{-1} we get

$$S = U\begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}U^{-1}$$

Now

$$S^n = U\begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}U^{-1}U\begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}U^{-1} \qquad n \text{ times}$$

which using Eq (6 11-19), becomes

$$S^n = U\begin{pmatrix} \lambda_1^n & 0 \\ 0 & \lambda_2^n \end{pmatrix}U^{-1} \tag{6 11-21}$$

order to obtain λ_1 and λ_2, we write

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}\begin{pmatrix} \xi \\ \eta \end{pmatrix} = \lambda\begin{pmatrix} \xi \\ \eta \end{pmatrix} \tag{6 11-22}$$

ıe above equation is called an eigenvalue equation, and can be rewrit-
n as

$$\begin{pmatrix} A - \lambda & B \\ C & D - \lambda \end{pmatrix}\begin{pmatrix} \xi \\ \eta \end{pmatrix} = 0 \tag{6 11-23}$$

ɔr nontrivial solutions to exist the determinant of the square matrix
ust vanish

$$\det\begin{pmatrix} A - \lambda & B \\ C & D - \lambda \end{pmatrix} = 0 \tag{6 11-24}$$

ıich gives us

$$(A - \lambda)(D - \lambda) - BC = 0$$

$$\lambda^2 - \lambda(A + D) + AD - BC = 0 \tag{6 11-25}$$

ıich using Eq (6 11-15) becomes

$$\lambda^2 - \lambda(A + D) + 1 = 0 \tag{6 11-26}$$

ıe above equation has two solutions, namely

$$\lambda_1 = \frac{A + D}{2} + \left[\left(\frac{A + D}{2}\right)^2 - 1\right]^{1/2} \tag{6 11-27a}$$

d

$$\lambda_2 = \frac{A + D}{2} - \left[\left(\frac{A + D}{2}\right)^2 - 1\right]^{1/2} \tag{6 11-27b}$$

ıese are the two eigenvalues of the system matrix S If we write

$$\cos\theta = \frac{A + D}{2} \tag{6 11-28}$$

ɜn we may write

$$\lambda_1 = \cos\theta + i\sin\theta = e^{i\theta} \tag{6 11-29}$$

d

$$\lambda_2 = \cos\theta - i\sin\theta = e^{-i\theta} \tag{6 11-30}$$

From Eqs (6 11-21), (6 11-29), and (6 11-30) it follows that if S^n should not diverge as n increases, θ must be real and hence

$$-1 < \frac{A + D}{2} < 1 \qquad (6\ 11\text{-}31)$$

This represents the stability condition of the resonator system

We now consider a resonator system formed by two mirrors of radii of curvatures R_1 and R_2 and separated by a distance d as shown in Fig 6 22 For a ray starting from the plane AB shown in Fig 6 22, the system matrix for one complete round trip would be

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} 1 - \frac{d}{R_1} - \frac{3d}{R_2} + \frac{2d^2}{R_1 R_2} & -\frac{2}{R_1} - \frac{2}{R_2} + \frac{4d}{R_1 R_2} \\ -\frac{d}{2} - \frac{d^2}{2R_1} + \left(1 - \frac{d}{R_2}\right)\left(\frac{3d}{2} - \frac{d^2}{R_1}\right) & -\frac{d}{R_1} + \left(1 - \frac{d}{R_2}\right)\left(1 - \frac{2d}{R_1}\right) \end{pmatrix} \qquad (6\ 11\text{-}32)$$

Comparing Eqs (6 11-16) and (6 11-32) we obtain

$$A = 1 - \frac{d}{R_1} - \frac{3d}{R_2} + \frac{2d^2}{R_1 R_2} \qquad (6\ 11\text{-}33)$$

and

$$D = 1 - \frac{3d}{R_1} - \frac{d}{R_2} + \frac{2d^2}{R_1 R_2} \qquad (6\ 11\text{-}34)$$

Substituting these in Eq (6 11-31) we obtain

$$-1 \leqslant 1 - \frac{2d}{R_1} - \frac{2d}{R_2} + \frac{2d^2}{R_1 R_2} \leqslant 1$$

or

$$0 \leqslant \left(1 - \frac{d}{R_1}\right)\left(1 - \frac{d}{R_2}\right) \leqslant 1 \qquad (6\ 11\text{-}35)$$

which is the stability condition [cf Eq (6 10-18)] Thus for a resonator to be stable R_1, R_2 and d must satisfy Eq (6 11-35)

Figure 6 23 depicts the stability diagram The shaded region corresponds to stable resonator configurations The origin corresponds to the

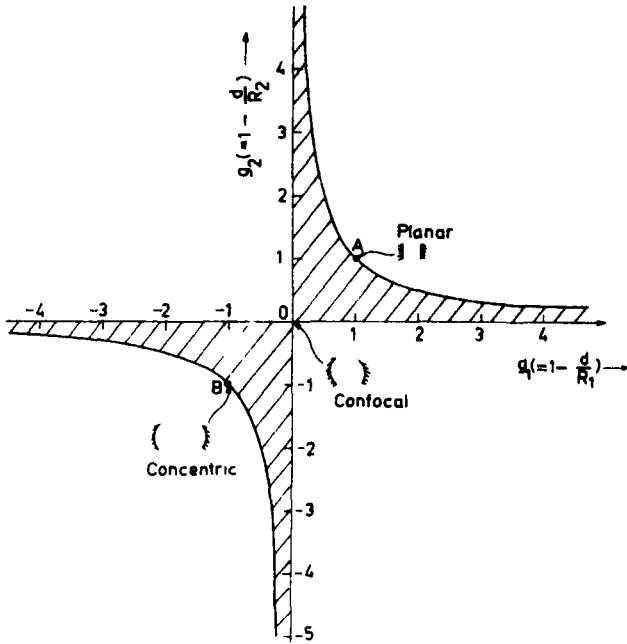



Fig 6 23 The stability diagram for optical resonators All resonator configurations which lie in the shaded region form stable configurations The origin corresponds to a confocal configuration A to a planar configuration and B to concentric configuration

confocal resonator The figure also shows explicitly the various resonator configurations It is interesting to note here that a resonator formed by two converging mirrors may become unstable if the power of the mirrors is too large Also a combination of a concave and a convex mirror may still form a stable resonator configuration

7

Vector Spaces and Linear Operators: Dirac Notation

7 1 Introduction

In this chapter we will introduce Dirac's bra and ket notation and also discuss the representation of observables by linear operators By imposing the communication relations we will solve the linear harmonic oscillator problem which will be used in the next chapter to study the quantized states of the radiation field Here we will discuss only those aspects of the bra and ket algebra which will be used later on For a thorough account, the reader is referred to the classic treatise by Dirac (1958a)

7 2 The Bra and Ket Notation†

A state of a system can be represented by a certain type of vector which we call a ket vector and represent by the symbol $|\,\rangle$ In order to distinguish the ket vectors corresponding to different states, we insert a label thus, the ket vector (or simply the ket) corresponding to the state A is described by the symbol $|A\rangle$ The kets form a linear vector space implying that if we have two states described by the kets $|A\rangle$ and $|B\rangle$, then the linear combination‡

$$C_1|A\rangle + C_2|B\rangle \tag{7 2-1}$$

is also a vector in the same space here C_1 and C_2 are arbitrary complex numbers Further to quote Dirac (p 16)

> each state of a dynamical system at a particular time corresponds to a ket vector the correspondence being such that if a state results from the superposition of certain other states its corresponding ket vector is expressible linearly in terms of the corresponding ket vectors of the other states and conversely

The conjugate imaginary of a ket vector $|A\rangle$ is denoted by $\langle A|$ and is called a bra vector (or, simply a bra) The scalar product of $|A\rangle$ and $\langle B|$ is

† The treatment will be based on the book by Dirac (1958a)

‡ It may be pointed out that $|P\rangle = C_1|A\rangle$ also represents the same state as $|A\rangle$ thus, if a ket is superposed on itself it corresponds to the same state

denoted by $\langle B \mid A\rangle$, which is a complex number Further

$$\langle B \mid A\rangle = \overline{\langle A \mid B\rangle} \tag{7 2-2}$$

where $\overline{\langle A \mid B\rangle}$ denotes the complex conjugate of the scalar product If we put $|B\rangle = |A\rangle$, we obtain from Eq (7 2-2) that $\langle A \mid A\rangle$ should be a real number We further impose that

$$\langle A \mid A\rangle \geqslant 0 \tag{7 2-3}$$

the equality sign holds if and only if $|A\rangle = 0$, i e , $|A\rangle$ is a null vector

The bra corresponding to $C|A\rangle$ is denoted by $C^*\langle A|$ where C is a complex number and C^* its complex conjugate A ket $|A\rangle$ is said to be normalized if

$$\langle A \mid A\rangle = 1 \tag{7 2-4}$$

and two kets are said to be orthogonal to each other if

$$\langle A \mid B\rangle = 0 \tag{7 2-5}$$

We may mention here the relationship between the Schrodinger wave functions developed in Chapter 2 to the bra and kets developed in this section If $|\Psi\rangle$ and $|\Phi\rangle$ represent the kets corresponding to the states described by the wave functions $\psi(\mathbf{r})$ and $\varphi(\mathbf{r})$ respectively, then

$$\langle \Phi \mid \Psi\rangle = \int \varphi^*(\mathbf{r})\psi(\mathbf{r})\,d\tau = \overline{\langle \Psi \mid \Phi\rangle} \tag{7 2-6}$$

7.3 Linear Operators

Let α be a linear operator which, acting on $|A\rangle$, produces $|B\rangle$

$$\alpha\,|A\rangle = |B\rangle \tag{7 3-1}$$

If $|B\rangle = 0$ for any $|A\rangle$, then $\alpha = 0$, i e , α is a null operator The operator is said to be linear if

$$\alpha(C_1\,|A\rangle + C_2\,|B\rangle) = C_1\alpha\,|A\rangle + C_2\alpha\,|B\rangle \tag{7 3-2}$$

where C_1 and C_2 are any complex numbers and $|A\rangle$ and $|B\rangle$ are arbitrary kets Two linear operators α and β are said to be equal if

$$\langle A|\,\alpha\,|A\rangle = \langle A|\,\beta\,|A\rangle \tag{7 3-3}$$

for any $|A\rangle$ The addition and multiplication of two linear operators α and β are defined through the equations

$$\left.\begin{aligned}(\alpha + \beta)\,|A\rangle &= \alpha\,|A\rangle + \beta\,|A\rangle\\ \{\alpha\beta\}\,|A\rangle &= \alpha\{\beta\,|A\rangle\}\end{aligned}\right\} \quad \text{for any } |A\rangle \tag{7 3-4}$$

An operator α acting on the bra $\langle A|$ results in $\langle A|\,\alpha$ and is defined through the equation

$$\{\langle A|\,\alpha\}\,|B\rangle = \langle A|\{\alpha\,|B\rangle\} \qquad \text{for any } |B\rangle \tag{7 3-5}$$

The adjoint of the operator α is denoted by $\bar{\alpha}$ and is defined through the relation that the conjugate imaginary of $\langle A|\alpha$ is $\bar{\alpha}|A\rangle$ Thus, if we put

$$\langle P| = \langle A|\alpha \tag{7 3-6}$$

then

$$|P\rangle = \bar{\alpha}|A\rangle \tag{7 3-7}$$

Hence

$$\langle B|\bar{\alpha}|A\rangle = \langle B|P\rangle = \overline{\langle P|B\rangle} = \overline{\langle A|\alpha|B\rangle} \tag{7 3-8}$$

If we now consider the operator adjoint of $\bar{\alpha}$, we obtain

$$\langle B|\bar{\bar{\alpha}}|A\rangle = \overline{\langle A|\bar{\alpha}|B\rangle} = \overline{\overline{\langle B|\alpha|A\rangle}} = \langle B|\alpha|A\rangle \tag{7 3-9}$$

because $\langle B|\alpha|A\rangle$ is just a complex number Since the above equation holds for arbitrary $\langle B|$ and $|A\rangle$ we must have

$$\bar{\bar{\alpha}} = \alpha \tag{7 3-10}$$

implying that the adjoint of the adjoint of an operator is the original operator itself If $\bar{\alpha} = \alpha$ then α is said to be a *real* or a *Hermitian* operator

We next consider two linear operators α and β whose adjoints are denoted by $\bar{\alpha}$ and $\bar{\beta}$, respectively Let

$$|P\rangle = \alpha\beta|A\rangle \tag{7 3-11}$$

then

$$\langle P| = \langle A|\overline{\alpha\beta} \tag{7 3-12}$$

Further, if $|Q\rangle = \beta|A\rangle$ then $|P\rangle = \alpha|Q\rangle$ and

$$\langle P| = \langle Q|\bar{\alpha} = \langle A|\bar{\beta}\bar{\alpha} \tag{7 3-13}$$

Thus

$$\overline{\alpha\beta} = \bar{\beta}\bar{\alpha} \tag{7 3-14}$$

and in general

$$\overline{\alpha\beta\gamma} = \bar{\gamma}\bar{\beta}\bar{\alpha} \tag{7 3-15}$$

It may be mentioned that the quantity

$$|P\rangle\langle Q| \tag{7 3-16}$$

is a linear operator, because when it acts on the ket $|A\rangle$ it produces the ket

$$C|P\rangle \tag{7 3-17}$$

where $C = \langle Q|A\rangle$ is a complex number However, operations like

$|P\rangle|Q\rangle$ or $\langle Q|\langle P|$ or $\alpha\langle Q|$ (where α is a linear operator) ar meaningless†

7 4. The Eigenvalue Equation

The equation

$$\alpha|P\rangle = a|P\rangle \qquad (7\ 4\text{-}$$

where a is a complex number, defines an eigenvalue equation, $|P\rangle$ is sai to be an eigenket of the operator α belonging to the eigenvalue a W premultiply Eq (7 4-1) by $\langle P|$ to obtain

$$\langle P|\alpha|P\rangle = a\langle P|P\rangle \qquad (7\ 4\text{-}$$

If we assume α to be a real operator, then by taking the comple conjugate of both sides we get

$$a^*\langle P|P\rangle = \overline{\langle P|\alpha|P\rangle} = \langle P|\bar{\alpha}|P\rangle = \langle P|\alpha|P\rangle = a\langle P|P\rangle \qquad (7\ 4\text{-}$$

where we have used Eq (7 3-8) and the fact that $\bar{\alpha} = \alpha$ Since $\langle P|P\rangle$ always positive (unless $|P\rangle$ is a null ket—which corresponds to the trivi solution) we obtain that a is a real number Thus, all eigenvalues of a re linear operator are real (cf Section 2 9)

We next show that eigenkets (of a real linear operator) belonging t different eigenvalues are necessarily orthogonal The conjugate of E (7 4-1) gives us

$$\langle P|\alpha = a^*\langle P| = a\langle P| \qquad (7\ 4\text{-}$$

because a is just a real number Now, if $|Q\rangle$ is another eigenket of belonging to the eigenvalue b then

$$\alpha|Q\rangle = b|Q\rangle \qquad (7\ 4\text{-}$$

Thus, postmultiplying Eq (7 4-4) by $|Q\rangle$ and premultiplying Eq (7 4- by $\langle P|$ we get

$$\langle P|\alpha|Q\rangle = a\langle P|Q\rangle = b\langle P|Q\rangle \qquad (7\ 4\text{-}$$

† The bras and kets follow the same axiomatic rules as row and column matrices and line operators follow the axiomatic rules of a square matrix Thus the multiplication of a ro matrix by a column matrix (cf $\langle B|A\rangle$) gives a number and multiplication of a colun matrix by a row matrix (cf $|A\rangle\langle B|$) gives a square matrix The multiplication of a squa matrix by a column matrix (cf $\alpha|P\rangle$) gives another column matrix, whereas multiplicatio of a square matrix by a row matrix (cf $\alpha\langle P|$) or of two column matrices (cf $|P\rangle|Q\rangle$) a meaningless

Since $b \neq a$, we get

$$\langle P \mid Q \rangle = 0 \tag{7.4-7}$$

implying that two eigenkets of a real linear operator (belonging to different eigenvalues) are orthogonal (cf Section 2 9) Further, if $|P_1\rangle$ and $|P_2\rangle$ are eigenkets of α belonging to the same eigenvalue a, i e ,

$$\alpha\,|P_1\rangle = a\,|P_1\rangle \quad \text{and} \quad \alpha\,|P_2\rangle = a\,|P_2\rangle$$

then

$$\alpha(C_1\,|P_1\rangle + C_2\,|P_2\rangle) = a(C_1\,|P_1\rangle + C_2\,|P_2\rangle)$$

implying that the ket $C_1\,|P_1\rangle + C_2\,|P_2\rangle$ is also an eigenket belonging to the *same* eigenvalue a Consequently, we can always choose appropriate linear combinations so that the eigenkets form an orthonormal set (cf the discussion in Section 2 9)

7 5 Observables

We assume that all measurable quantities (like position, momentum, energy etc) can be represented by linear operators Further, if one makes a precise measurement of such a quantity, one gets one of the eigenvalues of the corresponding linear operator The operator corresponding to the observable must be real because the result of measurement of any observable must be a real number, however, not every linear operator need correspond to an observable Finally, if a system is in a state described by $|P\rangle$ then the expectation value of an observable α is given by

$$\langle \alpha \rangle = \langle P|\,\alpha\,|P\rangle \tag{7.5-1}$$

The above equation may be compared with Eq (2 5-7) And, if $|P\rangle$ is such that

$$|P\rangle = C_1\,|\alpha_1\rangle + C_2\,|\alpha_2\rangle \tag{7.5-2}$$

where $|\alpha_1\rangle$ and $|\alpha_2\rangle$ are the normalized eigenkets of α belonging to the eigenvalues α_1 and α_2, respectively, then a measurement of the observable α on $|P\rangle$ would lead to either α_1 or α_2 with probabilities $|C_1|^2$ and $|C_2|^2$ respectively here we have assumed $|P\rangle$ to be normalized (i e , $\langle P \mid P \rangle = 1$) implying

$$|C_1|^2 + |C_2|^2 = 1 \tag{7.5-3}$$

In general, if the system is in any state then a measurement of α will

definitely lead to one of its eigenvalues Thus any ket $|P\rangle$ can be expressed as a linear combination of the eigenkets of the observable

$$|P\rangle = \sum_n C_n |\alpha_n\rangle \tag{7.5-4}$$

implying that the eigenkets of an observable form a complete set

7 6. The Harmonic Oscillator Problem

In Section 2 8 we solved the Schrodinger equation for the linear harmonic oscillator problem In this section we will use Dirac algebra to solve the harmonic oscillator problem for which the Hamiltonian is given by

$$H = \frac{p^2}{2m} + \tfrac{1}{2}m\omega^2 x^2 \tag{7 6-1}$$

Since H, p, and x are observables,

$$H = \bar{H}, \qquad p = \bar{p}, \qquad x = \bar{x} \tag{7 6-2}$$

Further, the operators x and p will be assumed to satisfy the commutation relation [cf Eq (2 6-3)]

$$[x, p] = xp - px = i\hbar \tag{7 6-3}$$

Our objective is to solve the eigenvalue equation

$$H|H'\rangle = H'|H'\rangle \tag{7 6-4}$$

where $|H'\rangle$ is the eigenket of the operator H belonging to the eigenvalue H'

It is convenient to introduce the dimensionless complex operator

$$a = \frac{1}{(2m\hbar\omega)^{1/2}}(m\omega x + ip) \tag{7 6-5}$$

The adjoint of a would be given by

$$\bar{a} = \frac{1}{(2m\hbar\omega)^{1/2}}(m\omega x - ip) \tag{7 6-6}$$

since x and p are observables [see Eq (7 6-2)] In terms of the above operators

$$\begin{aligned} \hbar\omega a\bar{a} &= \frac{1}{2m}(m\omega x + ip)(m\omega x - ip) \\ &= \frac{1}{2m}[m\omega^2 x^2 + p^2 - im\omega(xp - px)] \\ &= H + \tfrac{1}{2}\hbar\omega \end{aligned} \tag{7 6-7}$$

where we have used Eqs (7 6-1) and (7 6-3) Similarly

$$\hbar\omega\bar{a}a = H - \tfrac{1}{2}\hbar\omega \qquad (7\ 6\text{-}8)$$

Thus

$$H = \tfrac{1}{2}\hbar\omega(\bar{a}a + a\bar{a}) \qquad (7\ 6\text{-}9)$$

and

$$a\bar{a} - \bar{a}a = [a, \bar{a}] = 1 \qquad (7\ 6\text{-}10)$$

From Eq (7 6-7)

$$\hbar\omega a\bar{a}a = Ha + \tfrac{1}{2}\hbar\omega a \qquad (7\ 6\text{-}11)$$

and from Eq (7 6-8)

$$\hbar\omega a\bar{a}a = aH - \tfrac{1}{2}\hbar\omega a \qquad (7\ 6\text{-}12)$$

Thus

$$aH - Ha = [a, H] = \hbar\omega a \qquad (7\ 6\text{-}13)$$

Similarly

$$\bar{a}H - H\bar{a} = [\bar{a}, H] = -\hbar\omega\bar{a} \qquad (7\ 6\text{-}14)$$

Let

$$|P\rangle = a\,|H'\rangle$$

where $|H'\rangle$ is an eigenket of H belonging to the eigenvalue H' [see Eq (7 6-4)] Then

$$\begin{aligned}\hbar\omega\langle P\,|\,P\rangle &= \hbar\omega\langle H'|\,\bar{a}a\,|H'\rangle \\ &= \langle H'|\,H - \tfrac{1}{2}\hbar\omega\,|H'\rangle \qquad \text{[using Eq (7 6-8)]} \\ &= (H' - \tfrac{1}{2}\hbar\omega)\langle H'\,|\,H'\rangle \qquad \text{[using Eq (7 6-3)]}\end{aligned}$$

But $\langle P\,|\,P\rangle$ and $\langle H'\,|\,H'\rangle$ are positive numbers [see Eq (7 2-3)] and therefore

$$H' \geqslant \tfrac{1}{2}\hbar\omega \qquad (7\ 6\text{-}15)$$

and $H' = \tfrac{1}{2}\hbar\omega$ if and only if $|P\rangle = a\,|H'\rangle = 0$ (conversely, $a\,|H'\rangle$ is also a null ket only when $H' = \tfrac{1}{2}\hbar\omega$) That H' should be positive follows from Eq (7 6-1) and the fact that the expectation values of x^2 and p^2 should be positive or zero for any state of the system

Next let us consider the operator Ha operating on $|H'\rangle$

$$\begin{aligned}Ha\,|H'\rangle &= (aH - \hbar\omega a)\,|H'\rangle \qquad \text{[using Eq (7 6-13)]} \\ &= (aH' - \hbar\omega a)\,|H'\rangle \\ &= (H' - \hbar\omega)a\,|H'\rangle \qquad (7\ 6\text{-}16)\end{aligned}$$

Thus if $|H'\rangle$ is an eigenket of H then $a\,|H'\rangle$ is also an eigenket of H belonging to the eigenvalue $H' - \hbar\omega$ provided, of course, $a\,|H'\rangle$ is not a null ket, which will occur only if $H' = \frac{1}{2}\hbar\omega$ Thus if $H' \neq \frac{1}{2}\hbar\omega$, $H' - \hbar\omega$ is also an eigenvalue (provided $a\,|H'\rangle \neq 0$) Similarly if $H' - \hbar\omega \neq \frac{1}{2}\hbar\omega$ then $H' - 2\hbar\omega$ is also an eigenvalue of H We can thus say that $H' - \hbar\omega$, $H' - 2\hbar\omega$ are also eigenvalues provided $a\,|H'\rangle$, $aa\,|H'\rangle$, are not null kets This, however, cannot go on indefinitely because it will then contradict Eq (7 6-15) Further, it can terminate only at $H' = \frac{1}{2}\hbar\omega$ because then $a\,|\frac{1}{2}\hbar\omega\rangle = 0$

Now, using Eq (7 6-14)

$$\begin{aligned} H\bar{a}\,|H'\rangle &= (\bar{a}H + \hbar\omega\bar{a})\,|H'\rangle \\ &= (\bar{a}H' + \hbar\omega\bar{a})\,|H'\rangle \\ &= (H' + \hbar\omega)\bar{a}\,|H'\rangle \end{aligned} \tag{7 6-17}$$

implying that $(H' + \hbar\omega)$ is another eigenvalue of H, with $\bar{a}\,|H'\rangle$ as the eigenket belonging to it, unless $\bar{a}\,|H'\rangle = 0$ However, $\bar{a}\,|H'\rangle$ cannot be equal to zero, since it would lead to

$$\begin{aligned} 0 &= \hbar\omega a\bar{a}\,|H'\rangle = (H + \tfrac{1}{2}\hbar\omega)\,|H'\rangle \qquad \text{[using Eq (7 6-7)]} \\ &= (H' + \tfrac{1}{2}\hbar\omega)\,|H'\rangle \end{aligned}$$

giving $H' + \frac{1}{2}\hbar\omega = 0$ which contradicts Eq (7 6-15) Thus if H' is an eigenvalue then $H' + \hbar\omega$ is *always* another eigenvalue of H, and so are

$$H' + 2\hbar\omega \qquad H' + 3\hbar\omega,$$

and so on Hence the eigenvalues of the Hamiltonian for the linear harmonic oscillator problem are

$$\tfrac{1}{2}\hbar\omega \qquad \tfrac{3}{2}\hbar\omega, \qquad \tfrac{5}{2}\hbar\omega, \qquad \tfrac{7}{2}\hbar\omega, \tag{7 6-18}$$

extending to infinity which is the same as obtained in Section 2 8

We now relabel the eigenfunctions with the index n, thus $|n\rangle$ denotes the eigenfunction corresponding to the eigenvalues $(n+\frac{1}{2})\hbar\omega$

$$H\,|n\rangle = (n + \tfrac{1}{2})\hbar\omega\,|n\rangle \tag{7 6-19}$$

We assume that the states $|n\rangle$ are normalized so that

$$\langle m \mid n\rangle = \delta_{mn} \tag{7 6-20}$$

Further since $|0\rangle$ corresponds to $H' = \frac{1}{2}\hbar\omega$ we must have

$$a\,|0\rangle = 0 \tag{7 6-21}$$

Now for $n = 1\ 2\ 3,$ $\quad a\,|n\rangle$ is an eigenket of H belonging to the

eigenvalue $(n - \frac{1}{2})\hbar\omega$ [see Eq (7 6-16)], therefore $a\,|n\rangle$ must be a multiple of $|n - 1\rangle$

$$a\,|n\rangle = \alpha_n\,|n - 1\rangle \tag{7 6-22}$$

In order to determine α_n we calculate the square of the length of $a\,|n\rangle$

$$\langle n|\,\bar{a}a\,|n\rangle = |\alpha_n|^2\,\langle n - 1\,|\,n - 1\rangle = |\alpha_n|^2$$

But

$$\begin{aligned}\hbar\omega\,\langle n|\,\bar{a}a\,|n\rangle &= \langle n|\,(H - \tfrac{1}{2}\hbar\omega)\,|n\rangle \\ &= \langle n|\,(n + \tfrac{1}{2})\hbar\omega - \tfrac{1}{2}\hbar\omega\,|n\rangle \\ &= n\hbar\omega\,\langle n\,|\,n\rangle = n\hbar\omega\end{aligned}$$

Thus

$$|\alpha_n|^2 = n$$

and therefore

$$a\,|n\rangle = n^{1/2}\,|n - 1\rangle \tag{7 6-23}$$

Similarly

$$\bar{a}\,|n\rangle = (n + 1)^{1/2}\,|n + 1\rangle \tag{7 6-24}$$

Thus if $|0\rangle$ denotes the ground-state eigenket, then

$$|1\rangle = \frac{\bar{a}}{1^{1/2}}\,|0\rangle, \qquad |2\rangle = \frac{\bar{a}\bar{a}}{1^{1/2}\;2^{1/2}}\,|0\rangle = \frac{(\bar{a})^2}{(2!)^{1/2}}\,|0\rangle, \qquad |3\rangle = \frac{(\bar{a})^3}{(3!)^{1/2}}\,|0\rangle,$$

and, in general,

$$|n\rangle = \frac{(\bar{a})^n}{(n!)^{1/2}}\,|0\rangle \tag{7 6-25}$$

We can think of an excited state $|n\rangle$ of the oscillator as containing n quanta of energy $\hbar\omega$, in addition to the zero-point energy of $\frac{1}{2}\hbar\omega$ The operator $\bar{a}$, according to Eq (7 6-24), creates a quantum of energy, and, therefore, $\bar{a}$ is called the *creation* operator Similarly, the operator a, according to Eq (7 6-23), annihilates a quantum of energy, and, therefore, a is called the *annihilation* or *destruction* operator

7 6 1 The Number Operator

Consider the operator

$$N_{op} = \bar{a}a \tag{7 6 26}$$

Using Eq (7 6-8) we may write

$$H = (N_{op} + \tfrac{1}{2})\hbar\omega \tag{7 6-27}$$

Since

$$H\,|n\rangle = (n + \tfrac{1}{2})\hbar\omega\,|n\rangle \tag{7 6-28}$$

we have

$$\hbar\omega(N_{op} + \tfrac{1}{2})\,|n\rangle = (n + \tfrac{1}{2})\hbar\omega\,|n\rangle$$

or,

$$N_{op}\,|n\rangle = n\,|n\rangle \tag{7 6-29}$$

Thus $|n\rangle$ are also the eigenkets of N_{op}, the corresponding eigenvalue being n and since n takes the values 0, 1, 2   the operator N_{op} is called the number operator Obviously

$$\langle m|\,N_{op}\,|n\rangle = n\delta_{mn} \tag{7 6-30}$$

7 6 2 The Uncertainty Product

The quantities Δx and Δp which represent the uncertainties in x and p, are defined through the equations

$$\Delta x = (\langle x^2\rangle - \langle x\rangle^2)^{1/2} \tag{7 6-31}$$

$$\Delta p = (\langle p^2\rangle - \langle p\rangle^2)^{1/2} \tag{7 6-32}$$

where $\langle x^2\rangle$ represents the expectation value of x^2, etc We will calculate the expectation values of x, x^2, etc when the harmonic oscillator is in the state $|n\rangle$ Now

$$\langle n|\,x\,|n\rangle = \left(\frac{\hbar}{2m\omega}\right)^{1/2}\langle n|\,a + \bar{a}\,|n\rangle$$

[using Eqs (7 6-5) and (7 6-6)]

$$= \left(\frac{\hbar}{2m\omega}\right)^{1/2}[n^{1/2}\langle n\,|\,n - 1\rangle + (n + 1)^{1/2}\langle n\,|\,n + 1\rangle]$$

[using Eqs (7 6-23) and (7 6-24)]

$$= 0 \qquad \text{[using Eq (7 6-20)]} \tag{7 6-33}$$

$$\langle n|\,x^2\,|n\rangle = (\langle n|\,aa\,|n\rangle + \langle n|\,a\bar{a}\,|n\rangle + \langle n|\,\bar{a}a\,|n\rangle + \langle n|\,a\bar{a}\,|n\rangle)$$

$$= \frac{\hbar}{2m\omega}[0 + (n + 1) + n + 0]$$

$$= \frac{\hbar}{m\omega}(n + \tfrac{1}{2}) \tag{7 6-34}$$

Similarly

$$\langle n|p|n\rangle = 0 \tag{7.6-35}$$

and

$$\langle n|p^2|n\rangle = m\omega\hbar(n + \tfrac{1}{2}) \tag{7.6-36}$$

Thus

$$\Delta x\,\Delta p = (n + \tfrac{1}{2})\hbar \tag{7.6-37}$$

The minimum uncertainty product ($= \frac{1}{2}\hbar$) occurs for the ground state ($n = 0$). The result given by Eq. (7.6-37) is consistent with the uncertainty principle.

7.6.3 *The Coherent States*

Consider the eigenvalue equation

$$a|\alpha\rangle = \alpha|\alpha\rangle \tag{7.6-38}$$

where a is the annihilation operator defined through Eq. (7.6-5). The eigenkets defined by Eq. (7.6-38) are known as the coherent states.† In this section we will study some of the properties of the coherent states; these properties will be used in Chapter 8.

Since the eigenkets of an observable form a complete set, we expand $|\alpha\rangle$ in terms of the kets $|n\rangle$:

$$|\alpha\rangle = \sum_{n=0}^{\infty} C_n|n\rangle \tag{7.6-39}$$

Now

$$a|\alpha\rangle = \sum C_n a|n\rangle = \sum_{n=1} C_n n^{1/2}|n-1\rangle \tag{7.6-40}$$

Also

$$a|\alpha\rangle = \alpha|\alpha\rangle = \alpha\sum C_n|n\rangle \tag{7.6-41}$$

Thus

$$\alpha(C_0|0\rangle + C_1|1\rangle + \cdots) = C_1|0\rangle + C_2 2^{1/2}|1\rangle + C_3 3^{1/2}|2\rangle + \cdots$$

† In Section 8.4 we will show that when a laser is operated much beyond the threshold, it generates a coherent state excitation of the cavity mode. It is left as an exercise for the reader to show that the operator $a^\dagger$ cannot have any eigenkets and similarly a cannot have any eigenbras.

or

$$C_1 = \alpha C_0, \qquad C_2 = \frac{\alpha C_1}{2^{1/2}} = \frac{\alpha^2}{2^{1/2}} C_0$$
$$C_3 = \alpha \frac{C_2}{3^{1/2}} = \frac{\alpha^3}{(3!)^{1/2}} C_0, \tag{7.6-42}$$

In general,

$$C_n = \frac{\alpha^n}{(n!)^{1/2}} C_0 \tag{7.6-43}$$

Thus

$$|\alpha\rangle = C_0 \sum_n \frac{\alpha^n}{(n!)^{1/2}} |n\rangle \tag{7.6-44}$$

If we normalize $|\alpha\rangle$, we would get

$$1 = \langle \alpha \mid \alpha \rangle = |C_0|^2 \sum_n \sum_m \frac{\alpha^n \alpha^{*m}}{(n!)^{1/2}(m!)^{1/2}} \delta_{nm}$$
$$= |C_0|^2 \sum_n \frac{(|\alpha|^2)^n}{n!} = |C_0|^2 \exp(|\alpha|^2)$$

or

$$C_0 = \exp(-\tfrac{1}{2}|\alpha|^2) \tag{7.6-45}$$

within an arbitrary phase factor Substituting in Eq (7.6-44) we obtain

$$|\alpha\rangle = \exp(-\tfrac{1}{2}|\alpha|^2) \sum \frac{\alpha^n}{(n!)^{1/2}} |n\rangle \tag{7.6-46}$$

Notice that there is no restriction on the value of α, i e , α can take *any* complex value Further, if $|\beta\rangle$ is another eigenket of a belonging to the eigenvalue β, then

$$|\langle \alpha \mid \beta \rangle|^2 = \left| \exp(-\tfrac{1}{2}|\alpha|^2) \exp(-\tfrac{1}{2}|\beta|^2) \sum_n \sum_m \frac{\alpha^{*n}\beta^m}{(n!\,m!)^{1/2}} \langle n \mid m \rangle \right|^2$$
$$= \exp(-|\alpha|^2 - |\beta|^2) \left| \sum_n \frac{(\alpha^*\beta)^n}{n!} \right|^2$$
$$= \exp(-|\alpha|^2 - |\beta|^2 + \alpha^*\beta + \alpha\beta^*) = \exp(-|\alpha - \beta|^2) \tag{7.6-47}$$

Notice that the above equation implies that the eigenkets are not orthogonal (this is because a is not a real operator) they however become

approximately orthogonal for large values of $|\alpha - \beta|^2$ Further, the kets $|\alpha\rangle$ can be shown to satisfy the following relations

$$|\langle n \mid \alpha\rangle|^2 = \frac{1}{n!}|\alpha^2|^n \exp(-|\alpha|^2) \qquad (7\ 6\text{-}48)$$

$$\Delta x\, \Delta p = \tfrac{1}{2}\hbar \qquad (7\ 6\text{-}49)$$

where

$$\Delta x \equiv (\langle\alpha|\, x^2\, |\alpha\rangle - \langle\alpha|\, x\, |\alpha\rangle^2)^{1/2} \qquad (7\ 6\text{-}50)$$

and

$$\Delta p \equiv (\langle\alpha|\, p^2\, |\alpha\rangle - \langle\alpha|\, p\, |\alpha\rangle^2)^{1/2} \qquad (7\ 6\text{-}51)$$

Thus the uncertainty product $\Delta x\, \Delta p$ has the minimum value $(=\frac{1}{2}\hbar)$ for *all* coherent states Indeed it can be shown that if we solve the eigenvalue equation

$$a\psi = \frac{1}{(2m\hbar\omega)^{1/2}}(m\omega x + ip)\psi = \alpha\psi \qquad (7\ 6\text{-}52)$$

by replacing p by $-i\hbar\, d/dx$ [see Eq (2 3-8)] then the eigenfunctions would be displaced Gaussian functions and the uncertainty product is a minimum for the Gaussian function [see, e g , Ghatak and Lokanathan (1977)] It may be pointed out that the ground state ($n = 0$) wave function for the harmonic oscillator problem is also Gaussian [see Eq (2 8-25)]† and the corresponding uncertainty product is a minimum [see Eq (7 6-37) with $n = 0$]

7 7 Time Development of States

Let us consider a system described by the *time-independent* Hamiltonian H Let $|n\rangle$ represent the eigenkets of H belonging to the eigenvalue E_n

$$H\,|n\rangle = E_n\,|n\rangle \qquad (7\ 7\text{-}1)$$

Now, since the eigenkets form a complete set we may express an arbitrary ket $|\psi\rangle$ as a linear combination of $|n\rangle$

$$|\psi\rangle = \sum_n C_n\,|n\rangle \qquad (7\ 7\text{-}2)$$

The constants C_n can be obtained by premultiplying the above equation

† $H_0(\xi) = 1$

by $\langle m|$

$$\langle m \mid \psi\rangle = \sum C_n \langle m \mid n\rangle = \sum C_n \delta_{mn} = C_m \tag{7.7-3}$$

where we have used the orthonormality condition satisfied by the eigenkets. Thus

$$|\psi\rangle = \sum |n\rangle\langle n \mid \psi\rangle \tag{7.7-4}$$

which may be rewritten as

$$|\psi\rangle = \left\{\sum_n |n\rangle\langle n|\right\}|\psi\rangle \tag{7.7-5}$$

Since the above equation is valid for an arbitrary ket, the quantity inside the curly brackets must be the unit operator

$$\sum_n |n\rangle\langle n| = 1 \tag{7.7-6}$$

Equation (7.7-6) is often referred to as the completeness condition.

Next we are interested in finding out how a system (described by the time-independent Hamiltonian H) will evolve with time if the state of the system at $t = 0$ is described by the ket $|\psi(0)\rangle$. Now the ket $|\psi(t)\rangle$ (which describes the evolution of the system with time) would satisfy the time-dependent Schrodinger equation [Eq. (2.3-16)]

$$i\hbar \frac{d}{dt}|\Psi(t)\rangle = H|\Psi(t)\rangle \tag{7.7-7}$$

Since the Hamiltonian is independent of time, we can integrate the above equation to obtain

$$|\Psi(t)\rangle = e^{-iHt/\hbar}|\Psi(0)\rangle \tag{7.7-8}$$

where the exponential of an operator is *defined* through the power series

$$e^{O} = 1 + O + \frac{1}{2!}OO + \frac{1}{3!}OOO + \cdots \tag{7.7-9}$$

That Eq. (7.7-8) is the solution of Eq. (7.7-7) can be immediately seen by direct substitution. Using Eq. (7.7-5), we obtain

$$|\Psi(t)\rangle = e^{-iHt/\hbar}\sum_n |n\rangle\langle n \mid \Psi(0)\rangle \tag{7.7-10}$$

Replacing $e^{-iHt/\hbar}$ by the power series and using Eq. (7.7-1), we get

$$|\Psi(t)\rangle = \sum e^{-iE_n t/\hbar}|n\rangle\langle n \mid \Psi(0)\rangle \tag{7.7-11}$$

We may use the above equation to study the time development of coherent states Thus

$$|\Psi(0)\rangle = |\alpha\rangle = \exp(-\tfrac{1}{2}|\alpha|^2)\sum_n \frac{\alpha^n}{(n!)^{1/2}}|n\rangle \tag{7 7-12}$$

[see Eq (7 6-46)] Thus

$$\langle n \mid \Psi(0)\rangle = \langle n \mid \alpha\rangle = \exp(-\tfrac{1}{2}|\alpha|^2)\frac{\alpha^n}{(n!)^{1/2}} \tag{7 7-13}$$

Hence

$$|\Psi(t)\rangle = \exp(-\tfrac{1}{2}|\alpha|^2)\sum_n \left[\frac{\alpha^n}{(n!)^{1/2}}\right]\exp[-i(n+\tfrac{1}{2})\omega t]|n\rangle \tag{7 7-14}$$

where we have used the fact that

$$E_n = (n+\tfrac{1}{2})\hbar\omega \tag{7 7-15}$$

7 8 The Density Operator

Let $|0\rangle\ |1\rangle\ |2\rangle$ form a complete set of orthonormal kets, i e ,

$$\langle n \mid m\rangle = \delta_{mn} \tag{7 8-1}$$

and

$$\sum_n |n\rangle\langle n| = 1 \tag{7 8-2}$$

see Eqs (7 6-20) and (7 7-6) An arbitrary ket can be expanded in terms of $|n\rangle$

$$|P\rangle = \sum_n C_n|n\rangle, \qquad C_n = \langle n \mid P\rangle \tag{7 8-3}$$

A state can be characterized by the density operator ρ defined by the following equation

$$\rho = |P\rangle\langle P| \tag{7 8-4}$$

The trace† of an operator is defined to be equal to the sum of the diagonal matrix elements for any complete set of states, thus

$$\operatorname{Tr}\mathcal{O} = \sum_n \langle n|\mathcal{O}|n\rangle \tag{7 8-5}$$

† Abbreviated as Tr

where $\mathcal{O}$ is an arbitrary operator and

$$\mathcal{O}_{nm} \equiv \langle n|\mathcal{O}|m\rangle \quad ($$

is known as the (nm)th matrix element of the operator $\mathcal{O}$ with resp
kets $|0\rangle, |1\rangle$ etc being the base states the $n = m$ terms represe
diagonal elements Now

$$\mathrm{Tr}(|P\rangle\langle Q|) = \sum_n \langle n \mid P\rangle\langle Q \mid n\rangle$$

$$= \sum_n \langle Q \mid n\rangle\langle n \mid P\rangle$$

[because $\langle n \mid P\rangle$ and $\langle Q \mid n\rangle$ are complex nur

$$= \langle Q| \left\{\sum_n |n\rangle\langle n|\right\} |P\rangle \quad ($$

or

$$\mathrm{Tr}(|P\rangle\langle Q|) = \langle Q \mid P\rangle \quad ($$

where in the last step we have used Eq (7 8-2) Thus

$$\mathrm{Tr}\,\rho = \mathrm{Tr}\,|P\rangle\langle P| = \langle P \mid P\rangle = 1 \quad ($$

where we have assumed $|P\rangle$ to be normalized Further the expec
value of the operator $\mathcal{O}$ (when the system is in the state $|P\rangle$) is given

$$\langle\mathcal{O}\rangle = \langle P|\mathcal{O}|P\rangle = \sum_n \langle P|\mathcal{O}|n\rangle\langle n \mid P\rangle \quad \text{[using Eq (7 8-2)]}$$

$$= \sum \langle n \mid P\rangle\langle P|\mathcal{O}|n\rangle \quad \text{[because } \langle n \mid P\rangle \text{ is a number]}$$

$$= \sum \langle n|\rho\mathcal{O}|n\rangle \quad \text{[using Eq (7 8-4)]}$$

or

$$\langle\mathcal{O}\rangle = \mathrm{Tr}(\rho\mathcal{O}) \quad (7$$

Also

$$\rho_{nn} = \langle n \mid P\rangle\langle P \mid n\rangle = |C_n|^2 \quad (7$$

implying that the diagonal matrix elements of the density op
represent the probabilities of finding the system in the base states

Perhaps the most important application of the density operato
the field of statistical mechanics where we consider a large numb
identical systems, each system having a certain probability of bein
certain state If w_Ψ represents the probability of finding the system
state characterized by $|\Psi\rangle$, then the corresponding density opera

given by

$$\rho = \sum_{\Psi} w_{\Psi} |\Psi\rangle\langle\Psi| \tag{7 8-12}$$

where the summation is carried over all possible states of the system, the density operator contains all the information about the ensemble Since

$$\sum_{\Psi} w_{\Psi} = 1 \tag{7 8-13}$$

we obtain

$$\begin{aligned} \operatorname{Tr} \rho &= \sum_{\Psi} w_{\Psi} \operatorname{Tr} |\Psi\rangle\langle\Psi| \\ &= \sum_{\Psi} w_{\Psi} \langle\Psi | \Psi\rangle \\ &= \sum_{\Psi} w_{\Psi} = 1 \end{aligned} \tag{7 8-14}$$

Equation of Motion of the Density Operator

We calculate the time dependence of the density operator by differentiating Eq (7 8-12) with respect to time

$$i\hbar \frac{d\rho}{dt} = \sum_{\Psi} w_{\Psi} \left\{ i\hbar \frac{d|\Psi\rangle}{dt} \langle\Psi| + |\Psi\rangle \left[i\hbar \frac{d}{dt} \langle\Psi| \right] \right\} \tag{7 8-15}$$

But

$$i\hbar \frac{d}{dt} |\Psi\rangle = H |\Psi\rangle \tag{7 8-16}$$

and taking its conjugate imaginary

$$-i\hbar \frac{d}{dt} \langle\Psi| = \langle\Psi| H \tag{7 8-17}$$

Thus

$$\begin{aligned} i\hbar \frac{d\rho}{dt} &= \sum_{\Psi} w_{\Psi} [H |\Psi\rangle\langle\Psi| - |\Psi\rangle\langle\Psi| H] \\ &= H\rho - \rho H \end{aligned} \tag{7 8-18}$$

or

$$i\hbar \frac{d\rho}{dt} = -[\rho, H] \tag{7 8-19}$$

7.9 The Schrodinger and Heisenberg Pictures

While solving the linear harmonic oscillator problem in Section 7.6 we had assumed the observables x, p and H to be real operators and independent of time. This is the so-called Schrodinger picture and the time development of the ket describing the quantum mechanical system is obtained by solving the time-dependent Schrodinger equation

$$i\hbar \frac{d}{dt} |\Psi(t)\rangle = H |\Psi(t)\rangle \tag{7.9-1}$$

If the Hamiltonian is independent of time then we can integrate the above equation to obtain (see Section 7.7)

$$|\Psi(t)\rangle = e^{-iHt/\hbar} |\Psi(0)\rangle \tag{7.9-2}$$

where

$$e^{-iHt/\hbar} \equiv 1 - \left(\frac{iHt}{\hbar}\right) + \frac{1}{2!}\left(\frac{iHt}{\hbar}\right)\left(\frac{iHt}{\hbar}\right) + \cdots \tag{7.9-3}$$

Now, the expectation value of an observable characterized by the operator $\mathcal{O}$ is given by

$$\langle \mathcal{O} \rangle = \langle \Psi(t) | \mathcal{O} | \Psi(t) \rangle \tag{7.9-4}$$

Next, let $|n\rangle$ represent the eigenkets of the Hamiltonian H belonging to the eigenvalue E_n, i.e.

$$H |n\rangle = E_n |n\rangle \tag{7.9-5}$$

then [see Eq. (7.7-11)]

$$|\Psi(t)\rangle = \sum e^{-iE_n t/\hbar} |n\rangle [\langle n \mid \Psi(0)\rangle] \tag{7.9-6}$$

Thus in the Schrodinger picture, we may visualize the basis vectors (here $|n\rangle$) as a fixed set of vectors and $|\Psi(t)\rangle$ (describing the system) as moving.

Now, if we substitute for $|\Psi(t)\rangle$ from Eq. (7.9-6) in Eq. (7.9-4) we would get

$$\begin{aligned} \langle \mathcal{O} \rangle &= \langle \Psi(0) | e^{+iHt/\hbar} \mathcal{O} e^{-iHt/\hbar} | \Psi(0) \rangle \\ &= \langle \Psi(0) | \mathcal{O}_H(t) | \Psi(0) \rangle \end{aligned} \tag{7.9-7}$$

where the operator $\mathcal{O}_H(t)$ is defined by the following equation

$$\mathcal{O}_H(t) = e^{iHt/\hbar} \mathcal{O} e^{-iHt/\hbar} \tag{7.9-8}$$

Equations (7.9-4) and (7.9-7) tell us that the expectation values remain

the same if we endow the operator $\mathcal{O}_H(t)$ with the entire time dependence but assume that the kets are time independent This is known as the Heisenberg picture (and hence the subscript H) in which operators representing the observables change with time but the ket describing the state of the system is time independent From Eq (7 9-8), we have

$$\frac{d\mathcal{O}_H}{dt} = e^{+iHt/\hbar}\frac{\partial \mathcal{O}}{\partial t}e^{-iHt/\hbar} + \frac{i}{\hbar}e^{iHt/\hbar}[H\mathcal{O} - \mathcal{O}H]e^{-iHt/\hbar} \quad (7\ 9\text{-}9)$$

where the first term on the right-hand side allows for any explicit time dependence of the operator If there is no such explicit time dependence, we may write

$$i\hbar\frac{d\mathcal{O}_H(t)}{dt} = [e^{iHt/\hbar}\mathcal{O}e^{-iHt/\hbar}]H - H[e^{iHt/\hbar}\mathcal{O}e^{-iHt/\hbar}] \quad (7\ 9\text{-}10)$$

$$= \mathcal{O}_H(t)H - H\mathcal{O}_H(t) \quad (7\ 9\text{-}11)$$

or

$$i\hbar\frac{d\mathcal{O}_H(t)}{dt} = [\mathcal{O}_H(t), H] \quad (7\ 9\text{-}12)$$

The above equation is known as the Heisenberg equation of motion and gives the time dependence of an operator in the Heisenberg picture

If the Hamiltonian is assumed to be independent of time in the Schrodinger representation, then it is also independent of time in the Heisenberg representation

$$H_H(t) = e^{iHt/\hbar}He^{-iHt/\hbar} = e^{iHt/\hbar}e^{-iHt/\hbar}H = H \quad (7\ 9\text{-}13)$$

It should be mentioned that if H had an explicit time dependence, the analysis would have been much more involved [see, e g , Baym (1969) Chapter 5]

We next consider the operators α and β, which, in the Schrodinger representation, satisfy the commutation relation

$$[\alpha, \beta] = i\gamma \quad (7\ 9\text{-}14)$$

or

$$\alpha\beta - \beta\alpha = i\gamma \quad (7\ 9\text{-}15)$$

If we multiply on the left by $e^{iHt/\hbar}$ and on the right by $e^{-iHt/\hbar}$, we obtain

$$e^{iHt/\hbar}\alpha e^{-iHt/\hbar}e^{iHt/\hbar}\beta e^{-iHt/\hbar} - e^{iHt/\hbar}\beta e^{-iHt/\hbar}e^{iHt/\hbar}\alpha e^{-iHt/\hbar}$$
$$= ie^{iHt/\hbar}\gamma e^{-iHt/\hbar} \quad (7\ 9\text{-}16)$$

where we have inserted $e^{-iHt/\hbar}e^{iHt/\hbar}$ $(=1)$ between α and β Using Eq

(7 9-8), we get

$$[\alpha_H(t), \beta_H(t)] = i\gamma_H(t) \tag{7 9-17}$$

which shows the physical equivalence of Heisenberg and Schrödinger pictures

As an illustration we consider the harmonic oscillator problem However, before we do so, we note that

$$[x, p^n] = i\hbar n p^{n-1} = i\hbar \frac{\partial}{\partial p} p^n \tag{7 9-18}$$

and

$$[p, x^n] = -i\hbar n x^{n-1} = -i\hbar \frac{\partial}{\partial x} x^n \tag{7 9-19}$$

Thus if $P(p)$ and $X(x)$ can be expanded in a power series in p and x, respectively, we will have

$$[x, P(p)] = i\hbar \frac{\partial P}{\partial p} \tag{7 9-20}$$

and

$$[p, X(x)] = -i\hbar \frac{\partial X}{\partial x} \tag{7 9-21}$$

Now for the harmonic oscillator problem

$$H = \frac{p^2}{2m} + \tfrac{1}{2}m\omega^2 x^2 = H_H = \frac{p_H^2(t)}{2m} + \tfrac{1}{2}m\omega^2 x_H^2(t) \tag{7 9-22}$$

where we have used Eq (7 9-13) Thus, using Eq (7 9-12),

$$\frac{dx_H(t)}{dt} = \frac{1}{i\hbar}[x_H(t), H_H] = \frac{\partial H_H}{\partial p_H} = \frac{1}{m} p_H(t) \tag{7 9-23}$$

and

$$\frac{dp_H(t)}{dt} = \frac{1}{i\hbar}[p_H(t), H_H] = -\frac{\partial H_H}{\partial x_H} = -m\omega^2 x_H(t) \tag{7 9-24}$$

These are the Hamilton equations of motion (see, e g , Goldstein, 1950) The solutions of the above equations are

$$x_H(t) = x \cos \omega t + \frac{1}{m\omega} p \sin \omega t \tag{7 9-25}$$

$$p_H(t) = -m\omega x \sin \omega t + p \cos \omega t \tag{7 9-26}$$

where $x = x_H\ (t = 0)$ and $p = p_H\ (t = 0)$ represent the operators in the Schrodinger representation Further,

$$a_H(t) = \frac{1}{(2m\hbar\omega)^{1/2}}[m\omega x_H(t) + ip_H(t)] \qquad \text{[see Eq (7 6-5)]}$$

$$= ae^{-i\omega t} \tag{7 9-27}$$

where $a \equiv a_H\ (t = 0)$ Similarly

$$\bar{a}_H(t) = \bar{a}e^{+i\omega t} \tag{7 9-28}$$

where

$$\bar{a} \equiv \bar{a}_H\ (t = 0) \tag{7 9-29}$$

These relations will be used in the next chapter

8

Quantum Theory of Interaction of Radiation Field with Matter

8.1 Introduction

In this chapter we show that the electromagnetic field in a closed cavity can be considered as an infinite set of oscillators, each corresponding to a particular value of the wave vector and a particular direction of polarization. By imposing the commutation relations between the canonical variables, it is shown that the energy of each oscillator can increase or decrease by integral multiples of a certain quantum of energy, this quantum of energy is known as the *photon*. Having quantized the field, we show that the state which corresponds to a given number of photons for a particular mode does *not* correspond to the classical plane wave. Indeed, we show that the eigenstates of the annihilation operator (which are known as the coherent states) resemble the classical plane wave for large intensities.

In Section 3.3 we considered the interaction of the radiation field with matter using the semiclassical theory, i.e., we used a quantum mechanical description of the atom and a classical description of the electromagnetic field. In Section 8.5 we use the quantum mechanical description of the radiation field to study its interaction with an atom and thereby obtain explicit expressions for the Einstein A and B coefficients, which are shown to be identical to the one obtained in Chapter 3. We may mention here that the fully quantum mechanical theory automatically leads to spontaneous emissions which in the semiclassical theory had to be introduced in an ad hoc manner. Further, the theory also shows that the natural lineshape function is Lorentzian, this is explicitly shown in Appendix F.

In Section 8.6 we show that it is difficult to give a quantum mechanical description of the phase of the electromagnetic field.

8.2. Quantization of the Electromagnetic Field

We start with Maxwell's equations in free space

$$\nabla \times \mathscr{H} = \frac{\partial \mathscr{D}}{\partial t} = \epsilon_0 \frac{\partial \mathscr{E}}{\partial t} \tag{8.2-1}$$

$$\nabla \times \mathscr{E} = -\frac{\partial \mathscr{B}}{\partial t} = -\mu_0 \frac{\partial \mathscr{H}}{\partial t} \tag{8.2-2}$$

$$\nabla \cdot \mathscr{E} = 0 \tag{8.2-3}$$

$$\nabla \cdot \mathscr{B} = 0 \tag{8.2-4}$$

where we have assumed the absence of currents and free charges, $\mathscr{E}$, $\mathscr{D}$, $\mathscr{B}$, and $\mathscr{H}$ represent the electric field, electric displacement, magnetic induction, and magnetic field, respectively, ϵ_0 and μ_0 represent the permittivity and magnetic permeability of free space In writing Eqs (8.2-1)–(8.2-4), the MKS system of units has been used From Eq (8.2-4), it follows that $\mathscr{B}$ can be expressed as the curl of a vector

$$\mathscr{B} = \nabla \times \mathbf{A} = \mu_0 \mathscr{H} \tag{8.2-5}$$

where $\mathbf{A}$ is called the vector potential Thus Eqs (8.2-2) and (8.2-5) give us

$$\nabla \times \left(\mathscr{E} + \frac{\partial \mathbf{A}}{\partial t}\right) = 0 \tag{8.2-6}$$

Hence we may set

$$\mathscr{E} + \frac{\partial \mathbf{A}}{\partial t} = -\nabla \phi \tag{8.2-7}$$

or,

$$\mathscr{E} = -\nabla \phi - \frac{\partial \mathbf{A}}{\partial t} \tag{8.2-8}$$

where ϕ is known as the scalar potential Substituting Eq (8.2-8) into Eq (8.2-3), we get

$$\nabla^2 \phi + \frac{\partial}{\partial t}(\nabla \cdot \mathbf{A}) = 0 \tag{8.2-9}$$

Now, $\mathscr{B}$ is left unchanged if the gradient of any scalar quantity is added to $\mathbf{A}$

$$\mathbf{A} \rightarrow \mathbf{A}' = \mathbf{A} + \nabla \chi \tag{8.2-10}$$

where χ is any scalar function† We may choose χ such that

$$\nabla \cdot \mathbf{A} = 0 \tag{8 2-11}$$

This is known as the Coulomb gauge The scalar potential then satisfies the equation $\nabla^2\phi = 0$ and we may assume $\phi = 0$ Thus, we finally obtain

$$\mathcal{B} = \mu_0 \mathcal{H} = \nabla \times \mathbf{A} \tag{8 2-12}$$

and

$$\mathcal{E} = -\frac{\partial \mathbf{A}}{\partial t} \tag{8 2-13}$$

Substituting for $\mathcal{H}$ and $\mathcal{E}$ in Eq (8 2-1), we get

$$\nabla \times (\nabla \times \mathbf{A}) = -\epsilon_0\mu_0 \frac{\partial^2 \mathbf{A}}{\partial t^2} \tag{8 2-14}$$

If we now use the identity

$$\nabla \times (\nabla \times \mathbf{A}) = \nabla(\nabla \cdot \mathbf{A}) - \nabla^2 \mathbf{A} = -\nabla^2 \mathbf{A} \tag{8 2-15}$$

(because $\nabla \cdot \mathbf{A} = 0$), we finally obtain

$$\nabla^2 \mathbf{A} = \frac{1}{c^2}\frac{\partial^2 \mathbf{A}}{\partial t^2} \tag{8 2-16}$$

where $c\ [=(\epsilon_0\mu_0)^{-1/2}]$ represents the speed of light in free space Equation (8 2-16) represents the wave equation In order to solve the wave equation, we use the method of separation of variables

$$\mathbf{A}(\mathbf{r}, t) = \mathbf{A}(\mathbf{r})q(t) \tag{8 2-17}$$

Thus

$$q(t)\nabla^2 \mathbf{A}(\mathbf{r}) = \mathbf{A}(\mathbf{r})\frac{1}{c^2}\frac{d^2 q}{dt^2} \tag{8 2-18}$$

We next consider a Cartesian component (say the x component) of $\mathbf{A}(\mathbf{r})$ which we denote by $A_x(\mathbf{r})$, thus

$$\frac{c^2}{A_x(\mathbf{r})}\nabla^2 A_x(\mathbf{r}) = \frac{1}{q(t)}\frac{d^2 q}{dt^2} = -\omega^2 \qquad \text{(say)} \tag{8 2-19}$$

Thus

$$q(t) \sim e^{-i\omega t} \tag{8 2-20}$$

and

$$\nabla^2 A_x(\mathbf{r}) + k^2 A_x(\mathbf{r}) = 0 \tag{8 2-21}$$

† This is because $\nabla \times (\nabla\chi) = 0$ for arbitrary χ

where

$$k^2 \equiv \omega^2/c^2 \tag{8.2-22}$$

The solutions of Eq. (8.2-21) are plane waves, and similarly if we consider the y and z components, we obtain

$$\mathbf{A}(\mathbf{r}) = \hat{\mathbf{e}} e^{i\mathbf{k}\cdot\mathbf{r}} \tag{8.2-23}$$

where $\mathbf{k}\cdot\mathbf{k} = k^2$ and $\hat{\mathbf{e}}$ is the unit vector along $\mathbf{A}$. The condition $\nabla\cdot\mathbf{A} = 0$ gives us

$$\mathbf{k}\cdot\hat{\mathbf{e}} = 0 \tag{8.2-24}$$

implying that $\hat{\mathbf{e}}$ is at right angles to the direction of propagation $\mathbf{k}$; this is nothing but the transverse character of the wave; the vector $\hat{\mathbf{e}}$ denotes the polarization of the wave.

The allowed values of $\mathbf{k}$ (and hence of ω) are determined from the boundary conditions. If we assume the radiation to be confined in a cubical cavity of volume V $(= L^3)$ and use the periodic boundary condition, then

$$\mathbf{A}(x = 0, y, z) = \mathbf{A}(x = L, y, z), \text{ etc.} \tag{8.2-25}$$

giving

$$e^{ik_x L} = 1 = e^{ik_y L} = e^{ik_z L} \tag{8.2-26}$$

Thus

$$\left.\begin{aligned} k_x &= \frac{2\pi\nu_x}{L} \\ k_y &= \frac{2\pi\nu_y}{L} \\ k_z &= \frac{2\pi\nu}{L} \end{aligned}\right\} \nu_x, \nu_y, \nu_z = 0, \pm 1, \pm 2, \tag{8.2-27}$$

The complete solution of Eq. (8.2-16) is therefore given by

$$\mathbf{A}(\mathbf{r}, t) = \sum_{\lambda} [q_\lambda(t)\mathbf{A}_\lambda(\mathbf{r}) + q_\lambda^*(t)\mathbf{A}_\lambda^*(\mathbf{r})] \tag{8.2-28}$$

where

$$\mathbf{A}_\lambda(\mathbf{r}) = \hat{\mathbf{e}}_\lambda e^{i\mathbf{k}\cdot\mathbf{r}} \tag{8.2-29}$$

$$q_\lambda(t) = |q_\lambda| e^{-i\omega_\lambda t} \tag{8.2-30}$$

and the subscript λ signifies the various modes of the field [see Eq. (8.2-27)] including the two states of polarization. Thus a particular value

of λ corresponds to a particular set of values of ν_x, ν_y, ν_z and a particular direction of $\hat{\mathbf{e}}$ In Eq (8 2-28), the second term on the right-hand side is the complex conjugate of the first term, making $\mathbf{A}$ necessarily real Because of the allowed values of $\mathbf{k}_\lambda$ [see Eq (8 2-27)], we readily obtain

$$\iiint_V \mathbf{A}_\lambda \cdot \mathbf{A}_\mu^* \, d\tau = \iiint_V \mathbf{A}_\lambda \cdot \mathbf{A}_{-\mu} \, d\tau = V \delta_{\lambda\,\mu} \qquad (8\ 2\text{-}31)$$

where the integration is over the entire volume of the cavity

Using Eq (8 2-28), we obtain the following expressions for the electric and magnetic fields

$$\mathscr{E} = -\frac{\partial \mathbf{A}}{\partial t} = \sum_\lambda \mathscr{E}_\lambda \qquad (8\ 2\text{-}32)$$

$$\mathscr{H} = \frac{1}{\mu_0} \nabla \times \mathbf{A} = \sum_\lambda \mathscr{H}_\lambda \qquad (8\ 2\text{-}33)$$

where

$$\mathscr{E}_\lambda = i\omega_\lambda [q_\lambda(t) \mathbf{A}_\lambda(\mathbf{r}) - q_\lambda^*(t) \mathbf{A}_\lambda^*(\mathbf{r})] \qquad (8\ 2\text{-}34)$$

and

$$\mathscr{H}_\lambda = \frac{i}{\mu_0} \mathbf{k}_\lambda \times (q_\lambda \mathbf{A}_\lambda - q_\lambda^* \mathbf{A}_\lambda^*) \qquad (8\ 2\text{-}35)$$

The total energy of the radiation field is given by

$$H = \frac{1}{2} \int (\epsilon_0 \mathscr{E} \cdot \mathscr{E} + \mu_0 \mathscr{H} \cdot \mathscr{H}) \, d\tau \qquad (8\ 2\text{-}36)$$

Now

$$\begin{aligned}
\tfrac{1}{2}\epsilon_0 \int_V \mathscr{E} \cdot \mathscr{E} \, d\tau &= -\tfrac{1}{2}\epsilon_0 \int_V \sum_\lambda \sum_\mu \omega_\lambda \omega_\mu \Big[q_\lambda q_\mu \int \mathbf{A}_\lambda \cdot \mathbf{A}_\mu \, d\tau - q_\lambda q_\mu^* \int \mathbf{A}_\lambda \cdot \mathbf{A}_\mu^* \, d\tau \\
&\qquad - q_\lambda^* q_\mu \int \mathbf{A}_\lambda^* \cdot \mathbf{A}_\mu \, d\tau + q_\lambda^* q_\mu^* \int \mathbf{A}_\lambda^* \cdot \mathbf{A}_\mu^* \, d\tau \Big] \\
&= -\tfrac{1}{2}\epsilon_0 V \sum_\lambda \sum_\mu \omega_\lambda \omega_\mu [q_\lambda q_\mu \delta_{\lambda\,-\mu} - q_\lambda q_\mu^* \delta_{\lambda\,\mu} \\
&\qquad - q_\lambda^* q_\mu \delta_{\lambda\,\mu} + q_\lambda^* q_\mu^* \delta_{\lambda\,-\mu}] \\
&= -\tfrac{1}{2}\epsilon_0 V \sum_\lambda \omega_\lambda^2 [q_\lambda q_{-\lambda} + q_\lambda^* q_{-\lambda}^* - 2 q_\lambda q_\lambda^*]
\end{aligned} \qquad (8\ 2\text{-}37)$$

Similarly one can evaluate $\int \mathscr{H} \cdot \mathscr{H} \, d\tau$ The final result is

$$\tfrac{1}{2}\mu_0 \int \mathscr{H} \cdot \mathscr{H} \, d\tau = +\tfrac{1}{2}\epsilon_0 V \sum_\lambda \omega_\lambda^2 [q_\lambda q_{-\lambda} + q_\lambda^* q_{-\lambda}^* + 2 q_\lambda q_\lambda^*] \qquad (8\ 2\text{-}38)$$

where use has to be made of the vector identity

$$(\mathbf{a} \times \mathbf{b}) \cdot (\mathbf{c} \times \mathbf{d}) = (\mathbf{a} \cdot \mathbf{c})(\mathbf{b} \cdot \mathbf{d}) - (\mathbf{b} \cdot \mathbf{c})(\mathbf{a} \cdot \mathbf{d}) \qquad (8\,2\text{-}39)$$

and the relation

$$k_\lambda^2 = \frac{\omega_\lambda^2}{c^2} = \epsilon_0 \mu_0 \omega_\lambda^2 \qquad (8\,2\text{-}40)$$

Thus

$$H = 2\epsilon_0 V \sum_\lambda \omega_\lambda^2 q_\lambda(t) q_\lambda^*(t) \qquad (8\,2\text{-}41)$$

We next introduce the dimensionless variables $Q_\lambda(t)$ and $P_\lambda(t)$ defined through the equations

$$Q_\lambda(t) \equiv (\epsilon_0 V)^{1/2}[q_\lambda(t) + q_\lambda^*(t)] \qquad (8\,2\text{-}42)$$

and

$$P_\lambda(t) \equiv \frac{1}{i}(\epsilon_0 V \omega_\lambda^2)^{1/2}[q_\lambda(t) - q_\lambda^*(t)] \qquad (8\,2\text{-}43)$$

Thus

$$q_\lambda(t) = (4\epsilon_0 V \omega_\lambda^2)^{-1/2}[\omega_\lambda Q_\lambda(t) + iP_\lambda(t)] \qquad (8\,2\text{-}44)$$

$$q_\lambda^*(t) = (4\epsilon_0 V \omega_\lambda^2)^{-1/2}[\omega_\lambda Q_\lambda(t) - iP_\lambda(t)] \qquad (8\,2\text{-}45)$$

and

$$H = \sum_\lambda H_\lambda \qquad (8\,2\text{-}46)$$

where†

$$H_\lambda = \tfrac{1}{2}[P_\lambda^2 + \omega_\lambda^2 A_\lambda^2] \qquad (8\,2\text{-}47)$$

The Hamiltonian given by Eq (8 2-47) is identical to that of the linear harmonic oscillator [see Eq (7 6-1) with $m = 1$] which suggests that the

† Notice that

$$\frac{\partial H_\lambda}{\partial Q_\lambda} = \omega_\lambda^2 Q_\lambda = \omega_\lambda^2(\epsilon_0 V)^{1/2}(q_\lambda + q_\lambda^*) = i(\epsilon_0 V \omega_\lambda^2)^{1/2}(q_\lambda - q_\lambda^*) = -\dot{P}_\lambda$$

Similarly

$$\frac{\partial H_\lambda}{\partial P_\lambda} = \dot{Q}_\lambda$$

which are nothing but Hamilton's equations of motion (see e g , Goldstein, 1950) Thus Q_λ and P_λ are the canonical coordinates

electromagnetic field can be regarded as an infinite set of harmonic oscillators, one corresponding to each value of $\mathbf{k}_\lambda$ and to a particular direction of polarization

In order to quantize the electromagnetic field we use the same approach as in Section 7 6, we consider Q_λ and P_λ to be real operators satisfying the commutation relations [cf Eqs (7 6-3) and (7 9-17)]

$$[Q_\lambda(t), P_\lambda(t)] \equiv Q_\lambda(t)P_\lambda(t) - P_\lambda(t)Q_\lambda(t) = i\hbar \qquad (8\ 2\text{-}48)$$

$$[Q_\lambda(t), P_{\lambda'}(t)] = 0, \qquad \lambda \neq \lambda' \qquad (8\ 2\text{-}49)$$

$$[Q_\lambda(t), Q_\mu(t)] = 0 = [P_\lambda(t), P_\mu(t)] \qquad (8\ 2\text{-}50)$$

where all the operators are in the Heisenberg representation (see Section 7 9) We next introduce the dimensionless variables

$$a_\lambda(t) = (2\hbar\omega_\lambda)^{-1/2}[\omega_\lambda Q_\lambda(t) + iP_\lambda(t)] \qquad (8\ 2\text{-}51)$$

$$\bar{a}_\lambda(t) = (2\hbar\omega_\lambda)^{-1/2}[\omega_\lambda Q_\lambda(t) - iP_\lambda(t)] \qquad (8\ 2\text{-}52)$$

Since $[\omega_\lambda Q_\lambda(t) + iP_\lambda(t)]$ is proportional to $q_\lambda(t)$ [see Eq (8 2-44)] which has a time dependence of the form $e^{-i\omega_\lambda t}$, we may write [cf Eq (7 9-27)]

$$a_\lambda(t) = a_\lambda e^{-i\omega_\lambda t} \qquad (8\ 2\text{-}53)$$

Similarly

$$\bar{a}_\lambda(t) = \bar{a}_\lambda e^{i\omega_\lambda t} \qquad (8\ 2\text{-}54)$$

where

$$a_\lambda \equiv a_\lambda(0) \quad \text{and} \quad \bar{a}_\lambda \equiv \bar{a}_\lambda(0) \qquad (8\ 2\text{-}55)$$

Solving Eqs (8 2-51) and (8 2-52) for $Q_\lambda(t)$ and $P_\lambda(t)$ we obtain

$$Q_\lambda(t) = \left(\frac{\hbar}{2\omega_\lambda}\right)^{1/2}[\bar{a}_\lambda(t) + a_\lambda(t)] \qquad (8\ 2\text{-}56)$$

$$P_\lambda(t) = i\left(\frac{\hbar\omega_\lambda}{2}\right)^{1/2}[\bar{a}_\lambda(t) - a_\lambda(t)] \qquad (8\ 2\text{-}57)$$

Substituting the above expressions for $Q_\lambda(t)$ and $P_\lambda(t)$ in Eq (8 2-47), we obtain

$$\begin{aligned} H = \sum_\lambda H_\lambda &= \sum_\lambda \tfrac{1}{2}\hbar\omega_\lambda[\bar{a}_\lambda(t)a_\lambda(t) + a_\lambda(t)\bar{a}_\lambda(t)] \\ &= \sum_\lambda \tfrac{1}{2}\hbar\omega_\lambda[\bar{a}_\lambda a_\lambda + a_\lambda\bar{a}_\lambda] \end{aligned} \qquad (8\ 2\text{-}58)$$

If we now carry out an analysis similar to that followed in Section 7 6, we

obtain

$$(n_\lambda + \tfrac{1}{2})\hbar\omega_\lambda, \qquad n_\lambda = 0, 1, 2, \tag{8.2-59}$$

as the eigenvalues of H_λ and

$$\sum_\lambda (n_\lambda + \tfrac{1}{2})\hbar\omega_\lambda \tag{8 2-60}$$

as the eigenvalues of the total Hamiltonian H $(=\sum_\lambda H_\lambda)$ Thus, quantum mechanically, we can visualize the radiation field as consisting of an infinite number of simple harmonic oscillators, the energy of each oscillator can increase or decrease by integral multiples of $\hbar\omega_\lambda$ If we consider $\hbar\omega_\lambda$, as the energy of a *photon*, each oscillator can have energy corresponding to n_λ photons

The eigenkets of the total Hamiltonian would be

$$|n_1\rangle |n_2\rangle \quad |n_\lambda\rangle \quad = |n_1, n_2, \quad , n_\lambda, \quad \rangle \tag{8 2-61}$$

where n_λ represents the number of photons in the mode characterized by λ Thus

$$H\,|n_1, n_2, \quad , n_\lambda, \quad \rangle = \left[\sum_\lambda (n_\lambda + \tfrac{1}{2})\hbar\omega_\lambda\right] |n_1, n_2, \quad , n_\lambda, \quad \rangle \tag{8 2-62}$$

Further

$$a_\lambda\,|n_1, n_2, \quad , n_\lambda, \quad \rangle = (n_\lambda)^{1/2}\,|n_1, n_2, \quad , n_\lambda - 1, \quad \rangle \tag{8 2-63}$$

$$\bar{a}_\lambda\,|n_1, n_2, \quad , n_\lambda, \quad \rangle = (n_\lambda + 1)^{1/2}\,|n_1, n_2, \quad , n_\lambda + 1, \quad \rangle \tag{8 2-64}$$

[cf Eqs (7 6-23) and (7 6-24)], and

$$\langle n_1', n_2', \quad , n_\lambda', \quad |n_1, n_2, \quad , n_\lambda, \quad \rangle = \delta_{n_1 n_2}\,\delta_{n_2 n_2} \quad \delta_{n_\lambda n_\lambda} \tag{8 2-65}$$

Finally, the state of the radiation field need not be an eigenstate of H, it could be a superposition of the eigenstates like that given by the following equation

$$|\Psi\rangle = \sum_{n_1 n_2} C_{n_1 n_2 \quad n_\lambda}\,|n_1, n_2, \quad , n_\lambda, \quad \rangle \tag{8 2-66}$$

Physically $|C_{n_1 n_2}\ \ |^2$ would represent the probability of finding n_1 photons in the first mode, n_2 in the second mode, etc

8.3 The Eigenkets of the Hamiltonian

In this section we study the properties of the radiation field when it is in one of the eigenkets of the Hamiltonian In order to do so we have to

first express the electric field in terms of the operators a_λ and $\bar{a}_\lambda$ Now, using Eqs (8 2-44) and (8 2-51), we get

$$q_\lambda(t) = \left(\frac{\hbar}{2\epsilon_0 V\omega_\lambda}\right)^{1/2} a_\lambda(t) \tag{8 3-1}$$

which is now to be considered as an operator Using Eq (8 2-34), we may write

$$\mathscr{E}_\lambda(t) = i\left(\frac{\hbar\omega_\lambda}{2\epsilon_0 V}\right)^{1/2}[a_\lambda(t)e^{i\mathbf{k}_\lambda\cdot\mathbf{r}} - \bar{a}_\lambda(t)e^{-i\mathbf{k}_\lambda\cdot\mathbf{r}}]\hat{\mathbf{e}}_\lambda \tag{8 3-2}$$

where all the operators $[\mathscr{E}_\lambda(t),\ a_\lambda(t),$ and $\bar{a}_\lambda(t)]$ are in the Heisenberg representation (see Section 7 9) In the Schrödinger representation, we will have

$$\mathscr{E}_\lambda = i\left(\frac{\hbar\omega_\lambda}{2\epsilon_0 V}\right)^{1/2}(a_\lambda e^{i\mathbf{k}_\lambda\cdot\mathbf{r}} - \bar{a}_\lambda e^{-i\mathbf{k}_\lambda\cdot\mathbf{r}})\hat{\mathbf{e}}_\lambda \tag{8 3-3}$$

which will be independent of time We consider the state of the radiation field for which there are n_λ photons in the state λ The expectation value of $\mathscr{E}_\lambda$ in this state would be given by

$$\begin{aligned}\langle n_1, n_2, \ldots, n_\lambda, \ldots | \mathscr{E}_\lambda | n_1, n_2, \ldots, n_\lambda, \ldots\rangle \\ = \langle n_1 | n_1\rangle\langle n_2 | n_2\rangle \cdots \langle n_\lambda | \mathscr{E}_\lambda | n_\lambda\rangle \cdots = 0\end{aligned} \tag{8 3-4}$$

because

$$\langle n_\lambda | a_\lambda | n_\lambda\rangle = 0 = \langle n_\lambda | \bar{a}_\lambda | n_\lambda\rangle \tag{8 3-5}$$

Similarly

$$\begin{aligned}&\langle n_1, n_2, \ldots, n_\lambda, \ldots | \mathscr{E}_\lambda^2 | n_1, n_2, \ldots, n_\lambda, \ldots\rangle \\ &\quad = \langle n_1 | n_1\rangle\langle n_2 | n_2\rangle \cdots \langle n_\lambda | \mathscr{E}_\lambda^2 | n_\lambda\rangle \\ &\quad = -\frac{\hbar\omega_\lambda}{2\epsilon_0 V}\langle n_\lambda | (a_\lambda e^{i\mathbf{k}_\lambda\cdot\mathbf{r}} - \bar{a}_\lambda e^{-i\mathbf{k}_\lambda\cdot\mathbf{r}})(a_\lambda e^{i\mathbf{k}_\lambda\cdot\mathbf{r}} - \bar{a}_\lambda e^{-i\mathbf{k}_\lambda\cdot\mathbf{r}}) | n_\lambda\rangle \\ &\quad = \left(\frac{\hbar\omega_\lambda}{\epsilon_0 V}\right)(n_\lambda + \tfrac{1}{2})\end{aligned} \tag{8 3-6}$$

where use has been made of relations like (see Section 7 6)

$$\langle n_\lambda | a_\lambda\bar{a}_\lambda | n_\lambda\rangle = (n_\lambda + 1)^{1/2}\langle n_\lambda | a_\lambda | n_\lambda + 1\rangle = (n_\lambda + 1) \tag{8 3-7}$$

$$\langle n_\lambda | \bar{a}_\lambda a_\lambda | n_\lambda\rangle = (n_\lambda)^{1/2}\langle n_\lambda | \bar{a}_\lambda | n_\lambda - 1\rangle = n_\lambda \tag{8 3-8}$$

$$\langle n_\lambda | a_\lambda a_\lambda | n_\lambda\rangle = 0 \tag{8 3-9}$$

$$\langle n_\lambda | \bar{a}_\lambda\bar{a}_\lambda | n_\lambda\rangle = 0 \tag{8 3-10}$$

The uncertainty in $\mathscr{E}_\lambda$, $\Delta\mathscr{E}_\lambda$, can be defined through the relation

$$(\Delta\mathscr{E}_\lambda)^2 = \langle\mathscr{E}_\lambda^2\rangle - \langle\mathscr{E}_\lambda\rangle^2$$
$$= \frac{\hbar\omega_\lambda}{\epsilon_0 V}(n_\lambda + \tfrac{1}{2}) \qquad (8.3\text{-}11)$$

Equation (8.3-4) tells us that the expectation value of the electric field in the state $|n\rangle$ $(=|n_1, n_2, \ldots, n_\lambda, \ldots\rangle)$ is zero. Since the average of sine waves with random phases is zero, we may *loosely* say that the state $|n\rangle$ does not specify the phase.† Some authors tend to explain this by resorting to the uncertainty principle

$$\Delta E\, \Delta t \geqslant \hbar \qquad (8.3\text{-}12)$$

where ΔE is the uncertainty in the energy of the radiation field and Δt is related to the uncertainty in the phase angle through the relation

$$\Delta\phi = \omega\, \Delta t \qquad (8.3\text{-}13)$$

Since $E = (n + \tfrac{1}{2})\hbar\omega$, $\Delta E = \hbar\omega\, \Delta n$ and we obtain

$$\Delta n\, \Delta\phi \geqslant 1 \qquad (8.3\text{-}14)$$

If the number of photons is exactly known, then $\Delta n = 0$ and consequently there is no knowledge of the phase. However, such arguments are not rigorously correct because it is not possible to give a precise definition of $\Delta\phi$ (see Section 8.6). Nevertheless, we can say that the states described by $|n\rangle$ do not correspond to the classical electromagnetic wave with a certain phase.

Returning to Eq. (8.3-6), we notice that the states have an amplitude $(2\hbar\omega_\lambda/\epsilon_0 V)^{1/2}(n_\lambda + \tfrac{1}{2})^{1/2}$, for the mode λ, which is directly related to the number of photons.

8.4. The Coherent States

We next consider the radiation field to be in one of the coherent states which are the eigenkets of the operator a_λ (see Section 7.6). We will show that when the radiation field is in a coherent state, the field has properties very similar to that of a classical electromagnetic wave with a certain phase and amplitude. However, before we do so, we would like to discuss some of the properties of the coherent state.

† We say it loosely because it is not possible to define a phase operator which is real (see Section 8.6).

The coherent states satisfy the equation

$$a_\lambda |\alpha_\lambda\rangle = \alpha_\lambda |\alpha_\lambda\rangle \tag{8 4-1}$$

where α_λ, which represents the eigenvalue of a_λ, can be an arbitrary complex number In Section 7 6, we showed that

$$|\alpha_\lambda\rangle = \exp(-\tfrac{1}{2}|\alpha_\lambda|^2) \sum_{n_\lambda=0\,1} \frac{\alpha_\lambda^{n_\lambda}}{(n_\lambda!)^{1/2}} |n_\lambda\rangle \tag{8 4-2}$$

For convenience we drop the subscript λ so that the above equation becomes

$$|\alpha\rangle = \exp(-\tfrac{1}{2}|\alpha|^2) \sum_{n=0\,1\,2} \frac{\alpha^n}{(n!)^{1/2}} |n\rangle \tag{8 4-3}$$

Some of the important properties of $|\alpha\rangle$ are discussed below

(i) The expectation value of the number operator N_{op} ($= \bar{a}a$) is given by

$$\langle\alpha| N_{op} |\alpha\rangle = e^{-|\alpha|^2} \sum_m \frac{\alpha^{*m}}{(m!)^{1/2}} \langle m| \sum_n \frac{\alpha^n}{(n!)^{1/2}} n |n\rangle \qquad \text{[using Eq (7 6-29)]}$$

$$= e^{-|\alpha|^2} \sum_m \sum_n \frac{\alpha^{*m}\alpha^n}{(n!\, m!)^{1/2}} n\, \delta_{mn}$$

$$= e^{-|\alpha|^2} \sum_{n=1\,2} \frac{n\, |\alpha|^{2n}}{n!}$$

or

$$\langle\alpha| N_{op} |\alpha\rangle = e^{-|\alpha|^2} |\alpha|^2 \sum_{n=0\,1} \frac{|\alpha|^{2n}}{n!} = |\alpha|^2 = N \tag{8 4-4}$$

Thus the average number of photons (which we will denote by N) in the state $|\alpha\rangle$ is $|\alpha|^2$, and we may write

$$|\alpha\rangle = e^{-N/2} \sum_n \frac{\alpha^n}{n!} |n\rangle \tag{8 4-5}$$

(ii) From Eq (8 4-3) it readily follows that

$$|\langle n \mid \alpha\rangle|^2 = \frac{1}{n!} |\alpha|^{2n} \exp(-|\alpha|^2) = \frac{N^n e^{-N}}{n!} \tag{8 4-6}$$

Thus the probability of finding n photons in a coherent state is a Poisson distribution about the mean $|\alpha|^2$ [see also Eq (13 8–2)]

(iii) In Section 7 7, we showed that if the field is in the coherent state at $t = 0$, then at a later time t the state will be given by

$$|\Psi(t)\rangle = e^{-N/2} \sum_n \frac{\alpha^n}{(n!)^{1/2}} |n\rangle e^{-i(n+1/2)\omega t} \tag{8 4-7}$$

It is easy to see that

$$|\Psi(0)\rangle = e^{-N/2}\sum_n \frac{\alpha^n}{(n!)^{1/2}}|n\rangle = |\alpha\rangle \tag{8.4-8}$$

Further

$$\langle\Psi(t)|\, N_{op}\,|\Psi(t)\rangle = N \qquad \text{(independent of time)} \tag{8.4-9}$$

and†

$$\langle\Psi(t)|\,\tilde{a}\,|\Psi(t)\rangle = e^{-N}\sum_m\sum_n \frac{\alpha^{*m}\alpha^n}{(m!\,n!)^{1/2}}\, e^{i(m-n)\omega t}(n+1)^{1/2}\langle m\,|\,n+1\rangle$$

$$= e^{i\omega t}e^{-N}\sum \frac{\alpha^*\,|\alpha|^{2n}}{n!} = \alpha^* e^{i\omega t} \tag{8.4-10}$$

Similarly (or, taking the complex conjugate of the above equation)

$$\langle\Psi(t)|\,a\,|\Psi(t)\rangle = \alpha e^{-i\omega t} \tag{8.4-11}$$

We now consider the radiation field to be in the coherent state and calculate the expectation value of $\mathscr{E}$ and $\mathscr{E}^2$

$$\langle\Psi(t)|\,\mathscr{E}\,|\Psi(t)\rangle = i\left(\frac{\hbar\omega}{2\epsilon_0 V}\right)^{1/2}[\langle\Psi(t)|\,a\,|\Psi(t)\rangle e^{i\mathbf{k}\cdot\mathbf{r}} - \langle\Psi(t)|\,\tilde{a}\,|\Psi(t)\rangle e^{-i\mathbf{k}\cdot\mathbf{r}}]\hat{\mathbf{e}}$$

$$= i\left(\frac{\hbar\omega}{2\epsilon_0 V}\right)^{1/2}[\alpha e^{i(\mathbf{k}\cdot\mathbf{r}-\omega t)} - \alpha^* e^{-i(\mathbf{k}\cdot\mathbf{r}-\omega t)}]\hat{\mathbf{e}}$$

$$= -2\left(\frac{\hbar\omega}{2\epsilon_0 V}\right)^{1/2}|\alpha|\sin(\mathbf{k}\cdot\mathbf{r} - \omega t + \varphi)\hat{\mathbf{e}} \tag{8.4-12}$$

where

$$\alpha = |\alpha|\,e^{i\varphi} \tag{8.4-13}$$

Thus the coherent state can be interpreted to represent a harmonic wave with phase φ. In a similar manner, we can calculate the expectation value of $\mathscr{E}^2$. The result is

$$\langle\Psi(t)|\,\mathscr{E}\cdot\mathscr{E}\,|\Psi(t)\rangle = \left(\frac{\hbar\omega}{2\epsilon_0 V}\right)[4\,|\alpha|^2\sin^2(\mathbf{k}\cdot\mathbf{r} - \omega t + \varphi) + 1] \tag{8.4-14}$$

It may be noted that even for $\alpha = 0$ (which corresponds to the ground state‡ of the Hamiltonian), the expectation value of $\mathscr{E}^2$ is nonzero, this is

† In the Heisenberg representation, the expectation value of $\tilde{a}$ would have been $\langle\Psi(0)|\,\tilde{a}(t)\,|\Psi(0)\rangle = \langle\alpha|\,\tilde{a}e^{i\omega t}\,|\alpha\rangle = \alpha^* e^{i\omega t}$, which is the same as expressed by Eq. (8.4-10).

‡ The eigenket $|\alpha\rangle$ (of the operator a), corresponding to the eigenvalue 0, satisfies $a\,|\alpha\rangle = 0$—see Eq. (8.4-1). This eigenket is therefore $|0\rangle$—see Eq. (7.6-21).

called the zero-point fluctuation, which is responsible for spontaneous emissions Finally, the uncertainty in $\mathscr{E}$ would be given by [cf Eq (8 3-11)]

$$(\Delta\mathscr{E})^2 = \langle\Psi(t)|\mathscr{E}\cdot\mathscr{E}|\Psi(t)\rangle - \langle\Psi(t)|\mathscr{E}|\Psi(t)\rangle^2 \quad (8\,4\text{-}15)$$

$$= \left(\frac{\hbar\omega}{2\epsilon_0 V}\right) \quad (8\,4\text{-}16)$$

Notice that the uncertainty $\Delta\mathscr{E}$ is independent of the amplitude $|\alpha|$, thus, the greater the intensity of the beam, the greater will be the proximity of the radiation field (corresponding to the coherent state) to the classical plane wave

Earlier in this section, we had evaluated $\langle\alpha| N_{op} |\alpha\rangle$—see Eq (8 4-4), in a similar manner, we can calculate $\langle\alpha| N_{op}^2 |\alpha\rangle$ from which we obtain†

$$\Delta N = [\langle\alpha| N_{op}^2 |\alpha\rangle - \langle\alpha| N_{op} |\alpha\rangle^2]^{1/2} = N^{1/2} \quad (8\,4\text{-}17)$$

or

$$\frac{\Delta N}{N} = \frac{1}{N^{1/2}} \quad (8\,4\text{-}18)$$

implying that the fractional uncertainty in the average number of photons goes to zero with increase in intensity

8 5 Transition Rates

The Hamiltonian of a system consisting of an atom in a radiation field can be written as

$$\begin{aligned} H &= H_0 + H' \\ &= H_a + H_r + H' \end{aligned} \quad (8\,5\text{-}1)$$

where H_a represents the Hamiltonian of the atom, H_r the Hamiltonian corresponding to the pure radiation field [see Eq (8 2-58)] and H' represents the interaction between the atom and the radiation field We will consider H_0 $(= H_a + H_r)$ as the unperturbed Hamiltonian and H' will be considered as a perturbation which will be assumed to be of the form [see Eqs (8 2-32) and (8 3-3)]

$$\begin{aligned} H' &= -e\mathscr{E}\cdot\mathbf{r} \\ &= -e\sum_\lambda \mathscr{E}_\lambda\cdot\mathbf{r} \end{aligned} \quad (8\,5\text{-}2)$$

† It is of interest to point out that even in nuclear counting, the uncertainty in the actual count is $N^{1/2}$ (see, e g , Bleuler and Goldsmith, 1952)

Now, the eigenvalue equations for H_a and H_r are

$$H_a |\psi_i\rangle = E_i |\psi_i\rangle \tag{8.5-3}$$

and

$$H_r |n_1, n_2, \ldots, n_\lambda, \ldots\rangle = \left[\sum_\lambda (n_\lambda + \tfrac{1}{2})\hbar\omega_\lambda\right] |n_1, n_2, \ldots, n_\lambda, \ldots\rangle \tag{8.5-4}$$

where $|\psi_i\rangle$ and E_i represent, respectively, the eigenkets and energy eigenvalues of the isolated atom and $|n_1, n_2, \ldots, n_\lambda, \ldots\rangle$ represents the eigenket of the pure radiation field with $\sum_\lambda (n_\lambda + \frac{1}{2})\hbar\omega_\lambda$ representing the corresponding eigenvalue [see Eq. (8.2-62)]. Thus the eigenvalue equation for H_0 will be

$$H_0 |u_n\rangle = W_n |u_n\rangle \tag{8.5-5}$$

where

$$W_n = E_n + \sum_\lambda (n_\lambda + \tfrac{1}{2})\hbar\omega_\lambda \tag{8.5-6}$$

and

$$|u_n\rangle = |i\rangle |n_1, n_2, \ldots, n_\lambda, \ldots\rangle = |i, n_1, n_2, \ldots, n_\lambda, \ldots\rangle \tag{8.5-7}$$

represents the ket corresponding to the atom being in state $|i\rangle$ and the radiation being in the state $|n_1, n_2, \ldots, n_\lambda, \ldots\rangle$.

Now the Schrodinger equation for the system consisting of the atom and the radiation field is

$$i\hbar \frac{\partial}{\partial t} |\Psi\rangle = (H_0 + H') |\Psi\rangle \tag{8.5-8}$$

As in Section 3.3, the solution of the above equation can be written as a linear combination of the eigenkets of H_0 [cf. Eq. (3.3-8)]

$$|\Psi\rangle = \sum_n C_n(t) e^{-iW_n t/\hbar} |u_n\rangle \tag{8.5-9}$$

Substituting in Eq. (8.5-8), we obtain

$$i\hbar \sum_n \left[\frac{dC_n}{dt} - \frac{iW_n}{\hbar} C_n\right] e^{-iW_n t/\hbar} |u_n\rangle$$
$$= \sum_n C_n(t) W_n e^{-iW_n t/\hbar} |u_n\rangle + H' \sum_n C_n(t) e^{-iW_n t/\hbar} |u_n\rangle \tag{8.5-10}$$

where we have used Eq. (8.5-5). It is immediately seen that the second term on the left-hand side cancels exactly with the first term on the

right-hand side If we now multiply by $\langle u_m|$ on the left, we would get

$$i\hbar \frac{dC_m}{dt} = \sum_n \langle u_m| H' |u_n\rangle e^{i(W_m - W_n)t/\hbar} C_n(t) \qquad (8\ 5\text{-}11)$$

Now, using Eq (8 3-3) for $\mathscr{E}_\lambda$, we obtain for H'

$$H' = -ie \sum_\lambda \left(\frac{\hbar\omega_\lambda}{2\epsilon_0 V}\right)^{1/2} [a_\lambda e^{i\mathbf{k}_\lambda \cdot \mathbf{r}} - \bar{a}_\lambda e^{-i\mathbf{k}_\lambda \cdot \mathbf{r}}]\hat{e}_\lambda \cdot \mathbf{r} \qquad (8\ 5\text{-}12)$$

Now, because of the appearance of a_λ and $\bar{a}_\lambda$ in the expression for H', the various terms in $\langle u_m| H' |u_n\rangle$ will be nonzero only if the number of photons in $|u_m\rangle$ differs by unity from the number of photons in $|u_n\rangle$ If we write out completely the right-hand side of Eq (8 5-11) it will lead to a coupled set of an infinite number of equations which would be impossible to solve We employ the perturbation theory and consider the absorption of one photon (of energy $\hbar\omega_i$) from the ith mode Further, if we assume the frequency ω_i to be very close to the resonant frequency corresponding to the transition from the atomic state $|a\rangle$ to $|b\rangle$, then Eq (8 5-11) reduces to the following two coupled equations

$$i\hbar \frac{dC_1}{dt} = H'_{12} e^{i(W_1 - W_2)t/\hbar} C_2(t) \qquad (8\ 5\text{-}13)$$

$$i\hbar \frac{dC_2}{dt} = H'_{21} e^{-i(W_1 - W_2)t/\hbar} C_1(t) \qquad (8\ 5\text{-}14)$$

where

$$|1\rangle = |a, n_1, n_2, \quad , n_\lambda, \quad , n_i, \quad \rangle \qquad (8\ 5\text{-}15)$$

and

$$|2\rangle = |b, n_1, n_2, \quad , n_\lambda, \quad , n_i - 1, \quad \rangle \qquad (8\ 5\text{-}16)$$

represent the initial and final states of the system Obviously, because of relations like Eqs (8 3-5),

$$\langle 1| H' |1\rangle = H'_{11} = 0 = H'_{22}$$

Further,

$$W_1 = E_a + \sum_{\substack{\lambda \\ \lambda \neq i}} (n_\lambda + \tfrac{1}{2})\hbar\omega_\lambda + (n_i + \tfrac{1}{2})\hbar\omega_i \qquad (8\ 5\text{-}17)$$

and

$$W_2 = E_b + \sum_{\substack{\lambda \\ \lambda \neq i}} (n_\lambda + \tfrac{1}{2})\hbar\omega_\lambda + (n_i - 1 + \tfrac{1}{2})\hbar\omega_i \qquad (8\ 5\text{-}18)$$

Thus,

$$W_1 - W_2 = (E_a - E_b) + \hbar\omega_i \tag{8.5-19}$$

Now,

$$\begin{aligned} H_{21}'^{*} = H_{12}' &= \sum_{\lambda} \langle a, n_1, n_2, \ldots, n_i, \ldots | (-ie)\left(\frac{\hbar\omega_\lambda}{2\epsilon_0 V}\right)^{1/2} \\ &\quad \times [a_\lambda e^{i\mathbf{k}_\lambda\cdot\mathbf{r}} - \bar{a}_\lambda e^{-i\mathbf{k}_\lambda\cdot\mathbf{r}}]\hat{\mathbf{e}}_\lambda \cdot \mathbf{r} | b, n_1, n_2, \ldots, n_i - 1, \ldots\rangle \\ &= \langle a, n_1, n_2, \ldots, n_i, \ldots | \left[+ie\left(\frac{\hbar\omega_i}{2\epsilon_0 V}\right)^{1/2} e^{-i\mathbf{k}_\lambda\cdot\mathbf{r}}\hat{\mathbf{e}}_i \cdot \mathbf{r}(n_i)^{1/2}\right] \\ &\quad \times | b, n_1, n_2, \ldots, n_i, \ldots\rangle \\ &= ie\left(\frac{\hbar\omega_i}{2\epsilon_0 V}\right)^{1/2}(n_i)^{1/2}\langle a| e^{-i\mathbf{k}_\lambda\cdot\mathbf{r}}\mathbf{r} |b\rangle \cdot \hat{\mathbf{e}}_i \end{aligned} \tag{8.5-20}$$

where we have used Eqs. (8.2-63)–(8.2-65). In calculating the above matrix element between the atomic states a and b we note that the atomic wave functions are almost zero for $r \gtrsim 10^{-8}$ cm. On the other hand, since in the optical region, $|\mathbf{k}|$ $(=2\pi/\lambda) \approx 10^5\ \mathrm{cm}^{-1}$, in the domain of integration, in the evaluation of the matrix element $\mathbf{k}_\lambda \cdot \mathbf{r} \ll 1$. Thus we may replace $e^{-i\mathbf{k}_\lambda\cdot\mathbf{r}}$ by unity; this is known as the dipole approximation, and we obtain

$$H_{12}' = -i\left(\frac{\hbar\omega_i}{2\epsilon_0 V}\right)^{1/2}(n_i)^{1/2}\mathscr{D}_{ab} \tag{8.5-21}$$

where $\mathscr{D}_{ab}$ is defined through Eq. (3.3-13a). We next try to solve Eq. (8.5-13) and (8.5-14) by using a method similar to that employed in Section 3.4. We assume that at $t = 0$ the system is in the state represented by $|1\rangle$, i.e.,

$$C_1(0) = 1, \qquad C_2(0) = 0 \tag{8.5-22}$$

On working out the solution one obtains

$$|C_2(t)|^2 = \left(\frac{\tilde{\Omega}_0}{2}\right)^2\left[\frac{\sin(\tilde{\Omega}'t/2)}{\tilde{\Omega}'/2}\right]^2 \tag{8.5-23}$$

where

$$\tilde{\Omega}_0^2 = \frac{2\omega_i n_i \mathscr{D}_{ab}^2}{\hbar\epsilon_0 V} \tag{8.5-24}$$

$$\tilde{\Omega}' = \left[\left(\frac{E_b - E_a}{\hbar} - \omega_i\right)^2 + \tilde{\Omega}_0^2\right]^{1/2} \tag{8.5-25}$$

For $(\hat{\Omega}_0^2 t^2/\hbar^2) \ll 1$, we obtain

$$|C_2(t)|^2 \simeq \frac{\omega n}{2\hbar\epsilon_0 V} \mathcal{D}_{ab}^2 \left(\frac{\sin\{[(E_b - E_a)/\hbar - \omega](t/2)\}}{\frac{1}{2}[(E_b - E_a)/\hbar - \omega]} \right)^2 \tag{8 5-26}$$

where we have dropped the subscript i. Equation (8 5-26) is the same as Eq (3 3-18) provided we replace E_0^2 by $2\hbar\omega n/\epsilon_0 V$ (see footnote on p 46) Similarly, if we consider the emission process, we would obtain

$$|C_2(t)|^2 = \frac{\omega(n + 1)}{2\hbar\epsilon_0 V} \mathcal{D}_{ab}^2 \left(\frac{\sin\{[(E_b - E_a)/\hbar - \omega](t/2)\}}{\frac{1}{2}[(E_b - E_a)/\hbar - \omega]} \right)^2 \tag{8 5-27}$$

where the states $|1\rangle$ and $|1\rangle$ are now given by

$$|1\rangle = |b, n_1, n_2, \quad , n_i, \quad \rangle \tag{8 5-28}$$

$$|2\rangle = |a, n_1, n_2, \quad , n_i + 1, \quad \rangle \tag{8 5-29}$$

Notice the presence of the term† $(n + 1)$ in Eq (8 5-27) This implies that even if the number of photons were zero originally, the emission probability is finite The term proportional to n in Eq (8 5-27) gives the probability for induced or stimulated emission since the rate at which it occurs is proportional to the intensity of the applied radiation On the other hand, the second term, which is independent of n, gives the spontaneous emission rate into the mode Observe that the spontaneous emission probability into a particular mode is exactly the same as the stimulated emission probability caused by a single photon into the same mode, a fact which we have used in Section 4 4

We next calculate the probability per unit time for spontaneous emission of radiation If we consider the emission to be in the solid angle $d\Omega$ then the number of modes for which the photon frequency lies between ω and $\omega + d\omega$ is (see Appendix D)

$$N(\omega)\, d\omega\, d\Omega = \frac{V\omega^2\, d\omega}{8\pi^3 c^3}\, d\Omega \tag{8 5-30}$$

Thus, the total probability of emission in the solid angle $d\Omega$ would be given by

$$\begin{aligned} \Gamma &= \frac{|\mathcal{D}_{ab}|^2}{2\hbar\epsilon_0 V} \int \left\{ \frac{\sin[(\omega_{ba} - \omega)(t/2)]}{(\omega_{ba} - \omega)/2} \right\}^2 \omega \frac{V}{8\pi^3 c^3} \omega^2\, d\omega\, d\Omega \\ &\simeq \frac{|\mathcal{D}_{ab}|^2}{16\pi^3\hbar\epsilon_0 c^3}\, d\Omega\, \omega_{ba}^3 \int \left\{ \frac{\sin[(\omega_{ba} - \omega)(t/2)]}{(\omega_{ba} - \omega)/2} \right\}^2 d\omega \end{aligned} \tag{8 5-31}$$

† The appearance of the term $(n + 1)$ is because of the relation

$$\hat{a}^{\dagger} |n\rangle = (n + 1)^{1/2} |n + 1\rangle$$

where use has been made of the fact that the quantity inside the curly brackets is a sharply peaked function around $\omega = \omega_{ba}$. Carrying out the integration, using the fact that

$$\int_{-\infty}^{+\infty} \frac{\sin^2 x}{x^2}\,dx = \pi \tag{8.5-32}$$

we obtain

$$\Gamma = \frac{1}{2\pi}\left(\frac{e^2}{4\pi\epsilon_0 \hbar c}\right)\frac{\omega^3}{c^2}|\langle b|\mathbf{r}|a\rangle\cdot\hat{\mathbf{e}}|^2\, t\, d\Omega \tag{8.5-33}$$

Thus the transition rate is given by

$$w_{sp} = \frac{1}{2\pi}\left[\frac{e^2}{4\pi\epsilon_0 \hbar c}\right]\frac{\omega^3}{c^2}|\langle b|\mathbf{r}|a\rangle\cdot\hat{\mathbf{e}}|^2\, d\Omega \tag{8.5-34}$$

In order to calculate the total probability per unit time for the spontaneous emission to occur (the inverse of which will give the spontaneous lifetime of the state), we must sum over the two independent states of polarization and integrate over the solid angle. Assuming the direction of $\mathbf{k}$ to be along the z axis, we may choose $\mathscr{E}$ to be along the x or y axes. Thus, if we sum

$$|\langle b|\mathbf{r}|a\rangle\cdot\hat{\mathbf{e}}|^2$$

over the two independent states of polarization, we obtain

$$|\langle b|\mathbf{r}|a\rangle\cdot\hat{\mathbf{x}}|^2 + |\langle b|\mathbf{r}|a\rangle\cdot\hat{\mathbf{y}}|^2 = P_x^2 + P_y^2 = P^2\sin^2\theta \tag{8.5-35}$$

where $\mathbf{P} \equiv \langle b|\mathbf{r}|a\rangle$ and θ is the angle that $\mathbf{P}$ makes with the z axis. Thus in order to obtain the Einstein A coefficient (which represents the total probability per unit time for the spontaneous emission to occur), in Eq. (8.5-34), we replace $|\langle b|\mathbf{r}|a\rangle\cdot\hat{\mathbf{e}}|^2$ by $|\langle b|\mathbf{r}|a\rangle|^2\sin^2\theta$ and integrate over the solid angle $d\Omega$ to obtain

$$\begin{aligned} A &= \frac{1}{2\pi}\left[\frac{e^2}{4\pi\epsilon_0 \hbar c}\right]\frac{\omega^3}{c^2}|\langle b|\mathbf{r}|a\rangle|^2\int_0^{\pi}\int_0^{2\pi}\sin^2\theta\sin\theta\, d\theta\, d\varphi \\ &= \frac{4}{3}\left[\frac{e^2}{4\pi\epsilon_0 \hbar c}\right]\frac{\omega^3}{c^2}|\langle b|\mathbf{r}|a\rangle|^2 \end{aligned} \tag{8.5-36}$$

which is identical to Eq. (3.3-29).

8.6 The Phase Operator†

Classically, for the Hamiltonian given by

$$H = \frac{p^2}{2m} + \tfrac{1}{2}m\omega^2 x^2 \tag{8.6-1}$$

† A major part of the discussion in this section is based on a paper by Susskind and Glogower (1964).

the Hamilton equations of motion (see, e g , Goldstein, 1950)

$$x = \frac{\partial H}{\partial p} \quad \text{and} \quad -p = \frac{\partial H}{\partial x} \tag{8 6-2}$$

gives us

$$x = \frac{p}{m} \quad \text{and} \quad -p = m\omega^2 x^2 \tag{8 6-3}$$

Thus

$$x = \frac{1}{m} p = -\omega^2 x \tag{8 6-4}$$

or

$$x + \omega^2 x = 0 \tag{8 6-5}$$

giving

$$x = Ae^{+i\phi} + Ae^{-i\phi}, \qquad \phi = \omega t \tag{8 6-6}$$

and

$$p = mx = im\omega(Ae^{i\phi} - Ae^{-i\phi}) \tag{8 6-7}$$

where A has been assumed to be real In general, we should have written $\phi = \omega t + \alpha$, but with proper choice of the time origin, we can always choose $\alpha = 0$ In quantum mechanics

$$x = \left(\frac{2\hbar}{m\omega}\right)^{1/2} (\bar{a} + a) \tag{8 6-8}$$

and

$$p = im\omega\left(\frac{2\hbar}{m\omega}\right)^{1/2} (\bar{a} - a) \tag{8 6-9}$$

[see Eqs (7 6-5) and (7 6-6)] If we compare Eqs (8 6-8) and (8 6-9) with Eqs (8 6-6) and (8 6-7), we are tempted to define the phase operator ϕ through the equations

$$\bar{a} = Re^{i\phi} \quad \text{and} \quad a = e^{-i\phi}R \tag{8 6-10}$$

where R and ϕ are assumed to be Hermitian operators This is indeed what has been done by Dirac (1958b) and Heitler (1954), however, this definition leads to inconsistent results as will be shown later Now, the number operator [see Eq (7 6-26)] is given by

$$N_{op} = \bar{a}a = Re^{i\phi}e^{-i\phi}R = R^2 \tag{8 6-11}$$

Thus

$$R = N_{op}^{1/2} \tag{8.6-12}$$

where the square root operator is defined through the relation

$$N_{op} = N_{op}^{1/2} N_{op}^{1/2} \tag{8.6-13}$$

Now

$$[a, \bar{a}] = a\bar{a} - \bar{a}a = 1 \qquad \text{[see Eq. (7.6-10)]} \tag{8.6-14}$$

Thus

$$e^{-i\phi}RRe^{i\phi} - Re^{i\phi}e^{-i\phi}R = 1 \tag{8.6-15}$$

or, premultiplying by $e^{i\phi}$, we get

$$N_{op}e^{i\phi} - e^{i\phi}N_{op} = e^{i\phi} \tag{8.6-16}$$

Hence

$$[N_{op}, e^{i\phi}] = e^{i\phi} \tag{8.6-17}$$

The above equation is satisfied if ϕ and N_{op} satisfy the commutation relation†

$$[N_{op}, \phi] = N_{op}\phi - \phi N_{op} = -i \tag{8.6-18}$$

The above equation suggests the uncertainty relation‡

$$\Delta N \, \Delta\phi \geqslant 1 \tag{8.6-19}$$

From the above equation it follows that if the number of light quanta of a wave are given, the phase of this wave is entirely undetermined and vice versa (Heitler, 1954). However, the above definition of the phase operator (and hence the uncertainty relation) is not correct because the

† This can be shown by noting that repeated application of Eq. (8.6-18) gives

$$N_{op}\phi^{m} - \phi^{m}N_{op} = -im\phi^{m-1}$$

Now

$$[N_{op}, e^{i\phi}] = \sum_m \frac{i^m}{m!}[N_{op}, \phi^m] = \sum \frac{i^{m-1}\phi^{m-1}}{(m-1)!} = e^{i\phi}$$

‡ Equation (8.6-18) may be compared with the equation [see Eq. (7.6-3)]

$$[x, p_x] = xp_x - p_x x = i\hbar$$

from which one can derive the uncertainty relation

$$\Delta x \, \Delta p_x \geqslant \hbar$$

(see any text on quantum mechanics, e.g., Powell and Craseman, 1961)

definition leads to inconsistent results For example, let us consider the matrix element

$$\langle m|[N_{op}, \phi]|n\rangle = -i\langle m|n\rangle = -i\,\delta_{m,n} \qquad (8\,6\text{-}20)$$

But

$$\langle m|[N_{op}, \phi]|n\rangle = \langle m|N_{op}\phi - \phi N_{op}|n\rangle = (m-n)\langle m|\phi|n\rangle \qquad (8\,6\text{-}21)$$

Thus

$$(m-n)\langle m|\phi|n\rangle = -i\,\delta_{m\,n} \qquad (8\,6\text{-}22)$$

which is certainly an impossibility One can also show that $e^{-i\phi}e^{+i\phi}$ is not a unit operator Thus the definition of a Hermitian ϕ through Eq (8 6-10) leads to inconsistent results Susskind and Glogower (1964) define the phase operator through the equations

$$P_{\exp-} \equiv a(N_{op}+1)^{-1/2} \qquad (8\,6\text{-}23)$$

and

$$P_{\exp+} \equiv (N_{op}+1)^{-1/2}\bar{a} \qquad (8\,6\text{-}24)$$

The subscripts to P imply that in a certain limiting sense $P_{\exp-}$ and $P_{\exp+}$ behave as $e^{-i\phi}$ and $e^{+i\phi}$, respectively These operators are used to define trigonometric functions of phase

$$P_{\cos} = \tfrac{1}{2}(P_{\exp^+} + P_{\exp}) \qquad (8\,6\text{-}25)$$

$$P_{\sin} = \frac{1}{2i}(P_{\exp^+} - P_{\exp^-}) \qquad (8\,6\text{-}26)$$

Both $P_{\cos}$ and $P_{\sin}$ are Hermitian operators, however, they do not commute Indeed, it is the noncommuting nature of $P_{\cos}$ and $P_{\sin}$ which makes $e^{+i\phi}$ not unitary

Although we leave the discussion on the phase operator rather abruptly here, what we have shown is that the phase operator for an oscillator cannot exist We refer the reader to the works of Susskind and Glogower (1964) and of Loudon (1973), where they show that the operators $P_{\cos}$ and $P_{\sin}$ are observables and that in a certain limiting sense they do become the classical functions of phase Indeed these operators can be used to define uncertainty relations

9

Properties of Laser Beams and Types of Lasers

9.1 Introduction

So far we have discussed the physics behind laser operation. Now we examine the properties of laser beams and the types of lasers. Basically the light from a laser and from any normal source of light are both electromagnetic in nature but they differ drastically in the monochromaticity, directionality, and other more detailed properties; these are briefly discussed in Section 9.2. Because of these special properties of a laser beam, they can be put to many interesting applications; some of these applications are discussed in Part II.

In Sections 9.3–9.8 we briefly discuss the various important types of lasers.

9.2 Coherence Properties of Laser Light

In this section we introduce the concepts of temporal and spatial coherence and discuss the difference between light coming out of a laser and the light from an ordinary source like the incandescent lamp.

9.2.1 Temporal Coherence

In order to understand the concept of temporal coherence, we consider a Michelson interferometer arrangement as shown in Fig. 9.1. S represents an extended near-monochromatic source of light (like a sodium lamp), G represents a beam splitter, and M_1 and M_2 are two plane mirrors. The mirror M_2 is fixed while the mirror M_1 can be moved either towards or away from G. Light from the source S is incident on G

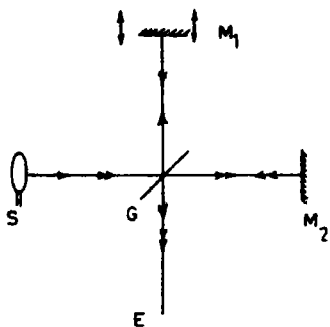



Fig. 9.1. The Michelson interferometer arrangement for studying the temporal coherence of light waves. S represents an extended source of light. G represents a beam splitter. M_1 and M_2 represent two plane mirrors; mirror M_2 is generally fixed and M_1 can be moved towards or away from G on a fine threaded screw. In general interference fringes of good contrast are visible from E when M_1 and M_2 are nearly equidistant from G while for sufficiently large distances of the mirror positions from G no interference fringes are visible.

and is divided into two portions: one part travels towards M_1 and is reflected back and the other part is reflected back from M_2. The two reflected waves interfere and produce interference fringes which are visible from E. When the mirrors M_1 and M_2 are nearly equidistant† from G, i.e., when the two waves traversing the two different paths take the same amount of time, then it is observed that the contrast of the interference fringes formed is good. If now the mirror M_1 is slowly moved away from G, then it is seen that for ordinary extended sources of light (like a sodium lamp) the contrast in the fringes goes on decreasing and when the difference between the distances from G to M_1 and M_2 is about a few millimeters to a few centimeters, the fringes are no longer visible. This decrease in contrast of the fringes can be explained as follows. The source S is emitting small wave trains of an average duration of τ_c (say) and there is no phase relationship between the different wave trains. This is in contrast to an infinitely long pure sinusoidal wave train which is also referred to as a monochromatic wave. When the difference in time taken by the wave trains to travel the paths G to M_1 and back and G to M_2 and back is much less than the average duration τ_c, then interference is produced between two wave trains each one being derived from the same wave train. Hence even though different wave trains emanating from the source S have no definite phase relationship, since one is superposing two wave trains derived from the same wave train, fringes of good contrast will be seen. On the other hand, if the difference in the time taken to traverse the path to M_1 and back and M_2 and back is much more than τ_c, then one is superposing two wave trains which are derived from two

† Assuming the beam splitter G to be coated on the back surface, it can easily be seen that the beam reflected from M_1 traverses G thrice whereas the beam reflected from M_2 traverses G only once. In order to compensate this additional path in the material of G, a compensating plate is introduced between G and M_2 (see, e.g., Jenkins and White, 1957). In the present discussion we are assuming G to be infinitesimally thin.

different wave trains, and since there is no definite phase relationship between two wave trains emanating from S, interference fringes will not be observed Hence as the mirror M_1 is moved, the contrast in the fringes becomes poorer and poorer and for large separations no fringes would be seen The time τ_c is referred to as the coherence time and the length of the wavetrain which is $c\tau_c$ (c being the velocity of light in free space) is referred to as the longitudinal coherence length It may be mentioned that there is no definite distance at which the interference pattern disappears, as the distance increases, the contrast of the fringes become gradually poorer and eventually the fringe system disappears

As an example, for the neon 6328-Å line from a discharge lamp, the interference fringes vanish if the path difference between the two mirrors is about a few centimeters Thus for this source, $\tau_c \sim 10^{-10}$ sec On the other hand, for the red cadmium line at $\lambda = 6438$ Å, the coherence length is about 30 cm, which gives $\tau_c \sim 10^{-9}$ sec

The decrease in the contrast of the fringes can also be interpreted as being due to the fact that the source S is not emitting at a single frequency but over a band of frequencies When the path difference is zero or very small, the different wavelength components produce fringes superimposed on one another and the fringe contrast is good On the other hand, when the path difference is increased, different wavelength components produce fringe patterns which are slightly displaced with respect to one another and the fringe contrast becomes poorer Thus the nonmonochromaticity of the light source can equally well be interpreted as the reason for the poor fringe visibility for large optical path differences

The equivalence of the above two approaches can also be seen by using Fourier analysis (see, e g , Section 3 5 and Appendix A) One can indeed show that a wave having a coherence time $\sim\tau_c$ is essentially a superposition of harmonic waves having frequencies in the region $\nu_0 - \Delta\nu/2 \leqslant \nu \leqslant \nu_0 + \Delta\nu/2$, where

$$\Delta\nu \sim 1/\tau_c \qquad (9\,2\text{-}1)$$

Thus, the longer the coherence time the smaller is the frequency width For an ordinary source, $\tau_c \sim 10^{-10}$ sec and

$$\Delta\nu \simeq 10^{10}\,\text{Hz} \qquad (9\,2\text{-}2)$$

For $\lambda = 6000$ Å, $\nu \simeq 5 \times 10^{14}$ Hz and

$$\frac{\Delta\nu}{\nu} \simeq \frac{1}{5 \times 10^4} = 0\,00002 \qquad (9\,2\text{-}3)$$

The quantity $\Delta\nu/\nu$ represents the monochromaticity and one can see that

even for an ordinary light source it is quite small In Section 3 5 we have discussed some of the mechanisms that lead to broadening of spectral lines emitted by an atom

In a laser, in contrast to an ordinary source of light, the optical resonant cavity is excited in different longitudinal modes of the cavity, which are specified by discrete frequencies of oscillation (A detailed discussion on modes of an optical resonator cavity has been given in Chapter 6) In an optical resonator without the laser medium, the finite loss of the resonator leads to an exponentially decaying output amplitude which leads to a finite linewidth of the output (see Section 6 3) On the other hand, in an actual laser oscillating in steady state, the loss is exactly compensated by the gain provided by the laser medium, and when the laser is oscillating in a single mode, the output is essentially a pure sinusoidal wave Superimposed on this are the random emissions arising out of spontaneous emissions, and it is this spontaneous emission which limits the ultimate monochromaticity of the laser (see Section 6 4)

In contrast to $\Delta\nu \sim 10^{10}$ Hz for an ordinary source of light for a well-controlled laser one can obtain $\Delta\nu \sim 500$ Hz, which gives $\tau_c \sim 2 \times 10^{-3}$ sec The corresponding coherence length is $c\tau_c \simeq 6 \times 10^7$ cm = 600 Km! Such long coherence lengths imply that the laser could be used for performing interference experiments with very large path differences

For a laser oscillating in many modes, the monochromaticity depends obviously on the number of oscillating modes Also for a pulsed laser, the minimum linewidth is limited by the duration of the pulse Thus, for a pulse of 10^{-9} sec duration from a mode-locked laser, $\tau_c \sim 10^{-9}$ sec and $\Delta\nu \sim 10^9$ Hz

9 2 2 Spatial Coherence

In order to understand the concept of spatial coherence, we consider the Young's double-hole experiment as shown in Fig 9 2 S represents a source placed in front of a screen with two holes S_1 and S_2, and the interference pattern between the waves emanating from S_1 and S_2 is observed on the screen T We restrict ourselves to the region near O for which the optical path lengths S_1O and S_2O are equal If S represents a point source then it illuminates the pinholes S_1 and S_2 with spherical waves Since the holes S_1 and S_2 are being illuminated coherently, the interference fringes formed near O will be of good contrast Consider now another point source $\tilde{S}$ placed near S and assume that the waves from S and $\tilde{S}$ have no phase relationship In such a case the interference pattern observed on the screen T will be a superposition of the intensity distributions of the interference patterns formed due to S and $\tilde{S}$ If $\tilde{S}$ is

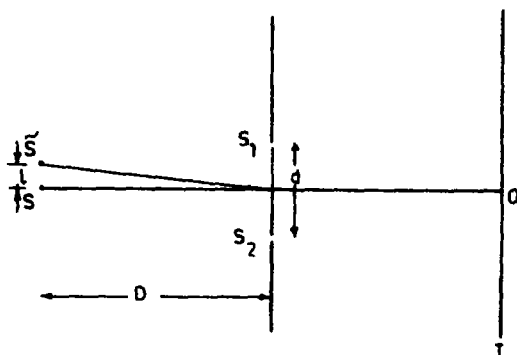



Fig 9 2 The Young's double-hole experiment for studying the spatial coherence of a light source S Interference fringes between light emanating from S_1 and S_2 are observed on a screen T If $S\tilde{S} = l \sim \lambda D/2d$ the interference pattern on the screen T is washed away

moved slowly away from S, the contrast of the fringes on T becomes poorer because of the fact that the interference pattern produced by $\tilde{S}$ is slightly shifted in relation to that produced by S For a particular separation, the interference maximum produced by S falls on the interference minimum produced by $\tilde{S}$ and the minimum produced by S falls on the maximum produced by $\tilde{S}$ For such a position the interference fringe pattern on the screen T is washed away

In order to obtain an approximate expression for the separation $S\tilde{S}$ for disappearance of fringes, we assume that S and O are equidistant from S_1 and S_2 If the position of $\tilde{S}$ is such that the path difference between $\tilde{S}S_2$ and $\tilde{S}S_1$ is $\lambda/2$ (where λ is the wavelength of the light used), then the source $\tilde{S}$ produces an interference minimum at O and the two fringe patterns would be out of step If we assume $S\tilde{S} = l$, $S_1S_2 = d$, and the distance between S and the plane of the pinholes as D, we obtain

$$\tilde{S}S_2 = \left[D^2 + \left(\frac{d}{2} + l\right)^2\right]^{1/2} \simeq D + \frac{1}{2D}\left(\frac{d}{2} + l\right)^2 \qquad (9\,2\text{-}4)$$

$$\tilde{S}S_1 = \left[D^2 + \left(\frac{d}{2} - l\right)^2\right]^{1/2} \simeq D + \frac{1}{2D}\left(\frac{d}{2} - l\right)^2 \qquad (9\,2\text{-}5)$$

where we have assumed that $D \gg d, l$ Thus for disappearance of fringes,

$$\tilde{S}S_2 - \tilde{S}S_1 = \frac{\lambda}{2} \simeq \frac{ld}{D}$$

or

$$l \simeq \lambda D/2d \qquad (9\,2\text{-}6)$$

For an extended source made up of independent point sources, one may say that good interference fringes will be observed as long as†

$$l \ll \lambda D/d \tag{9.2-7}$$

Equivalently, for a given source of width l, interference fringes of good contrast will be formed by interference of light from two points S_1 and S_2 separated by a distance d as long as

$$d \ll \lambda D/l \tag{9.2-8}$$

Since l/D is the angle subtended by the source at the slits (say θ) the above equation can also be written as

$$d \ll \lambda/\theta \tag{9.2-9}$$

The distance l_w ($= \lambda/\theta$) is referred to as the lateral coherence width. It can be seen from Eq. (9.2-9) that l_w depends inversely on θ.

As an example if we consider the sun as the source, then we have

$$\theta = 32' = \frac{\pi \times 32}{180 \times 60}\,\text{rad} \simeq 0.01\,\text{rad} \tag{9.2-10}$$

Thus, the lateral coherence width will be about

$$\frac{5 \times 10^{-5}}{10^{-2}} \simeq 0.005\,\text{cm}$$

where we have assumed that we have placed a filter to pass light around $\lambda = 5 \times 10^{-5}$ cm. Thus if we have a pair of pinholes separated by a distance much less than 0.005 cm and illuminated by the sun, interference fringes of good contrast can be obtained on a screen (see Fig. 9.3).

Using ordinary extended light sources, one must pass the light through a pinhole in order to produce a spatially coherent beam of light. In contrast, the laser beam is highly spatially coherent. For example, Fig. 9.4 shows the interference pattern obtained by Nelson and Collins (1961) by placing a pair of slits of width 0.00075 cm separated by a distance 0.0054 cm on the end of a ruby rod in the ruby laser. The interference pattern agrees with the theoretical calculation to within 20%. To show that the spatial coherence is indeed due to laser action, they showed that below threshold no interference pattern was observed, only a uniform darkening of the photographic plate was obtained.

† For an incoherent source of extension $\lambda D/d$ for every point on the source, there is a point at a distance $\lambda D/2d$ which produces fringes separated by half a fringe width. Thus for a source of such an extension the visibility of interference fringes would be very poor.

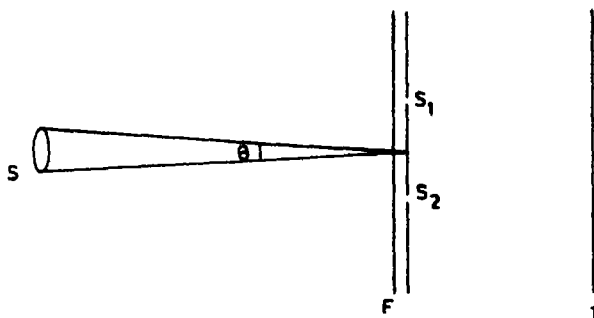



Fig 9 3 An extended source S illuminates a pair of pinholes S_1 and S_2 through a filter F (which passes light of wavelength λ), and interference fringes of good quality will be observed on the screen T as long as $d \ll \lambda/\theta$

9 2 3 Directionality

An ordinary source of light (like a sodium lamp or an incandescent source) radiates in all possible directions On the other hand, the output from a laser may be very close to an ideal uniform plane wave whose divergence is primarily due to diffraction effects For example, for the laser oscillating in its lowest mode in the cavity shown in Fig 9 5, the output coming out of the plane mirror would be approximately plane and the slow spreading is due to diffraction effects

We saw in Section 6 10 that the fundamental mode of a general spherical resonator is a Gaussian beam whose waist size and waist position can be obtained in terms of the parameters of the resonator

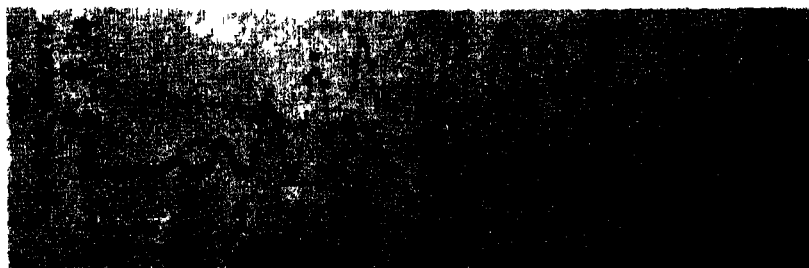



Fig 9 4 The double-slit interference pattern obtained by placing a pair of slits each 0 00075 cm wide and separated by a distance of 0 00541 cm across the diameter of a ruby rod (a) shows the actual interference pattern and (b) shows a densitometer trace of the interference pattern The dots correspond to a theoretical calculation assuming that a plane wave strikes the pair of slits After Nelson and Collins (1961) Photograph courtesy Dr D F Nelson

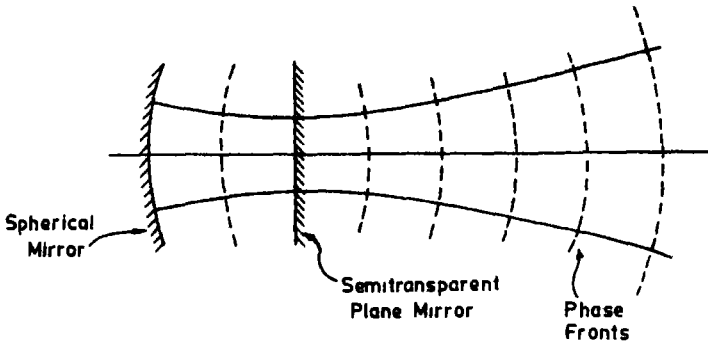



Fig 9 5 A resonator made up of a spherical mirror and a plane mirror The wave front of the output beam coming out of the plane mirror when the resonator is oscillating in the lowest-order mode is plane, and for suitable values of the resonator parameters the spreading will be small

system For simplicity, let us consider a resonator made up of two mirrors of the same radii of curvatures R and separated by a distance d The waist size is determined from Eq (6 10–15) and on substituting the value of $g_1 = g_2 = 1 - d/R$ it can be shown that

$$w_0^4 = \frac{\lambda^2 dR}{4\pi^2}\left(2 - \frac{d}{R}\right) \tag{9 2-11}$$

The value of w_0 in fact determines the diffraction divergence of the beam Thus using Eq (6 10-4) for the beam size at a distance z from the waist, we obtain for large values of z,

$$w(z) \simeq \lambda z/\pi w_0 \tag{9 2-12}$$

and the divergence semiangle can be written as (see Fig 9 6)

$$\theta \simeq w(z)/z = \lambda/\pi w_0 \tag{9 2-13}$$

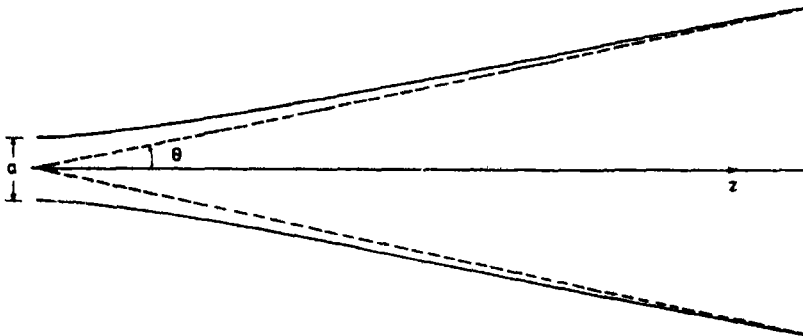



Fig 9 6 Diffraction of a beam limited by an aperture of width a

Thus, the larger the beam waist size, the smaller is the divergence Let us first consider a confocal system of mirrors of radii of curvatures equal to 100 cm and separated by a distance $d = 100$ cm Then for $\lambda = 0.5 \times 10^{-4}$ cm, one has

$$w_0 \simeq 0.03 \text{ cm}$$

and (9 2-14)

$$\theta \simeq 0.45 \times 10^{-3} \text{ rad} = 90 \text{ sec of arc}$$

On the other hand, for a resonator made up of two mirrors each having a radius of curvature equal to 500 m and separated by the same distance of 100 cm, one obtains

$$w_0 \simeq 0.5 \text{ cm}$$

and (9 2-15)

$$\theta \simeq 6 \text{ sec of arc}$$

Thus depending on the resonator parameters, the resulting Gaussian beam has a large or a small beam waist size and correspondingly a small or a large divergence In addition, the waist size can be increased by making use of a simple optical system consisting of a pair of convex lenses as shown in Fig 9 7

It may be of interest to note here that if we place a small filament of linear dimension Δx at the focal plane of a convex lens then because of the finite size of the filament, the angular divergence of the emerging beam will be given by (see Fig 9 8)

$$\Delta\theta \simeq \Delta x/f \qquad (9\ 2\text{-}16)$$

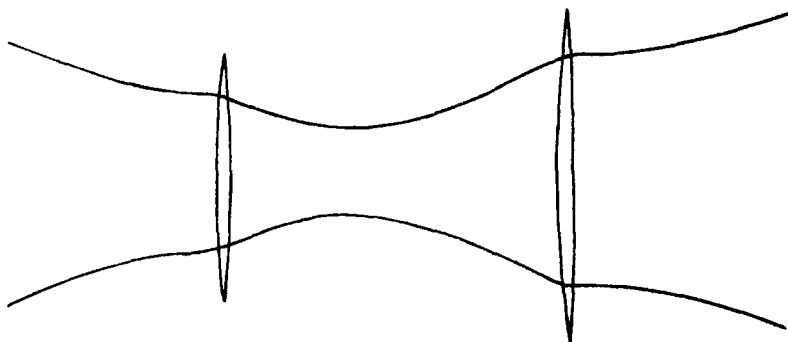



Fig 9 7 The waist size of the Gaussian beam can be increased by passing it through a pair of convex lenses

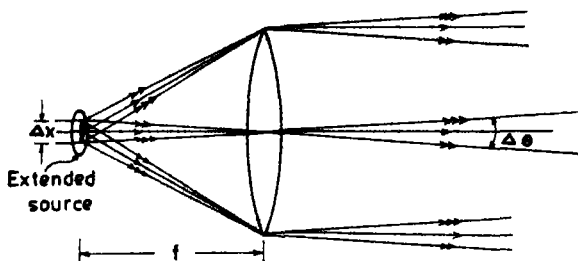



Fig 9 8 A filament of linear dimension Δx is placed on the focal plane of a convex lens of focal length f The angular divergence because of the finite size of the filament $\sim \Delta x/f$

For the emergent beam to be diffraction limited (i e , the divergence to be caused only by diffraction)

$$\Delta x/f \ll \lambda/a$$

or

$$\Delta x \ll \lambda f/a \tag{9 2-17}$$

For $\lambda \simeq 5 \times 10^{-5}$ cm, $f = 10$ cm, and $a \simeq 5$ cm, one obtains

$$\Delta x \ll 10^{-4} \text{ cm} \tag{9 2-18}$$

Such small filaments are indeed difficult to produce and also the power would be negligibly small

Because of the high directionality of the laser beam, it can be focused to extremely small dimensions It may be mentioned that according to geometrical optics, a plane wave is said to focus to a point by an aberrationless converging lens, however, if we take into account the effects of diffraction, one can show that the plane wave focuses to a region of radius of about $\lambda f/a$, where λ is the wavelength of light, f is the focal length, and a is the radius of the lens (or of the laser beam, whichever is smaller) For a typical lens, one may assume that $a \sim 5$ cm, $f \sim 10$ cm and for $\lambda \simeq 5 \times 10^{-5}$ cm, the area of the spot will be approximately

$$\pi\left(\frac{\lambda f}{a}\right)^2 \sim 3 \times 10^{-8} \text{ cm}^2 \tag{9 2-19}$$

Thus, a 10-W laser beam, will produce an intensity of about

$$3 \times 10^9 \text{ W/cm}^2$$

at the focal plane of the lens (see Fig 14 2) Such high intensities lead to numerous industrial applications of the laser†

† For an incoherent source, even under the most ideal conditions the brightness at the focal point can at the most be only equal to that of the source

9.3. The Ruby Laser

The first laser to be operated successfully was the ruby laser which was fabricated by Maiman in 1960 It consists of a single crystal of ruby whose ends are flat and one of which is completely silvered and the other partially silvered, the two ends thus form a resonant cavity Ruby consists of aluminium oxide with some of the aluminium atoms replaced by chromium For a good lasing action, about 0.05% of the atoms are that of chromium, however, higher concentrations of chromium have also been used The energy levels of the chromium ion are shown in Fig 9 9 The states in the bands labeled E_1 and E_2 have a very small lifetime ($\leq 10^{-9}$ sec) whereas the metastable state (labeled M) has a much longer lifetime (~milliseconds) The ruby crystal is placed inside a xenon flash lamp as shown in Fig 9 10 The flashlamp is connected to a capacitor

Fig 9 9 Energy levels of chromium ions in the ruby crystal The chromium ions absorb the radiation emitted by the flash lamp and get excited to levels in the energy bands E_1 and E_2 from where they suffer a fast nonradiative transition to the metastable level M Population inversion between M and G results in lasing action at $\lambda \sim 6943$ Å

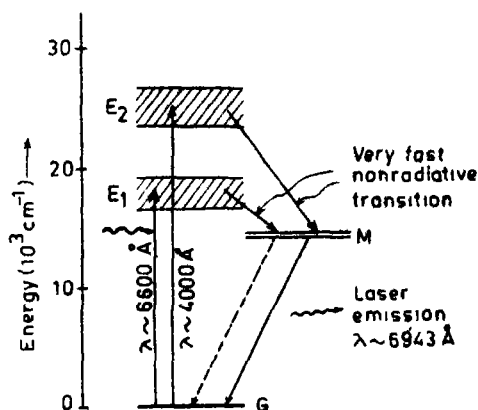
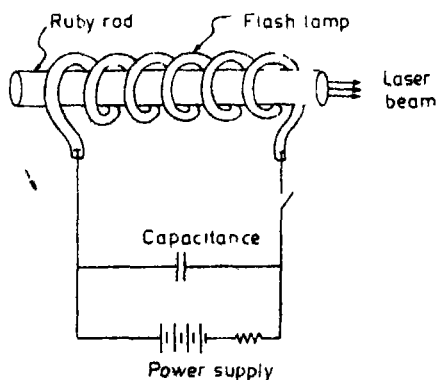



Fig 9 10 Schematic of a ruby laser The flashlamp is connected to a capacitor which discharges a few thousand joules of energy in a few milliseconds Some of the resulting light energy is absorbed by the chromium ions in ruby, leading to population inversion One end of the ruby rod is completely polished while the other is partially reflecting

which discharges a few thousand joules of energy in a few milliseconds This results in a power output (from the flash lamp) of a few megawatts, and a part of this energy is absorbed by the chromium ions (in the ground state), resulting in their excitation to an energy level inside the bands E_1 and E_2, transitions to E_1 and E_2 are caused by radiation of wavelength ~6600 Å and ~4000 Å, respectively The chromium ions make a very fast nonradiative transition from the excited state to the metastable state marked M and since this state has a long lifetime, the number of atoms in this state keeps on increasing, this eventually leads to a state of population inversion between states M and G Once a state of population inversion is achieved, lasing action is triggered by spontaneously emitted photons.

The flashlamp operation leads to a pulsed output of the laser The moment the flashlamp stops operating, the population in the upper level is depleted very fast and lasing action stops till the arrival of the next flash from the lamp Even in the short period of a few tens of microseconds in which the ruby is lasing, one finds that the emission is made up of spikes of high intensity emissions as shown in Fig 9 11 This phenomenon is termed spiking of the laser (see Section 4 6) and occurs due to the following mechanism As soon as the flashlamp power reaches the threshold level, the laser oscillation sets in and depletes most of the atoms in the upper level to a stage when the laser action ceases Thus the laser emission stops for a few microseconds, within which time the flashlamp again pumps the ground state atoms to the upper level and again laser oscillation begins This process repeats itself till the flashlamp power falls below the threshold value and the lasing action stops

Very high peak power pulses from the ruby laser can be obtained by Q switching (see Section 6 6)

In addition to the pumping scheme shown in Fig 9 10 using a helical flashlamp, one may employ other methods of pumping also Thus, for example, Fig 9 12 shows another optical pumping scheme in which the laser rod and the flash lamp coincide with the focal lines of a cylindrical reflector of an elliptical cross section The special property of the elliptical reflector is that all the energy emerging from one of its foci after

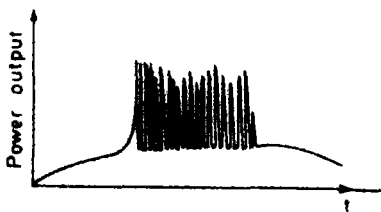



Fig 9 11 Characteristic spiking of a ruby laser

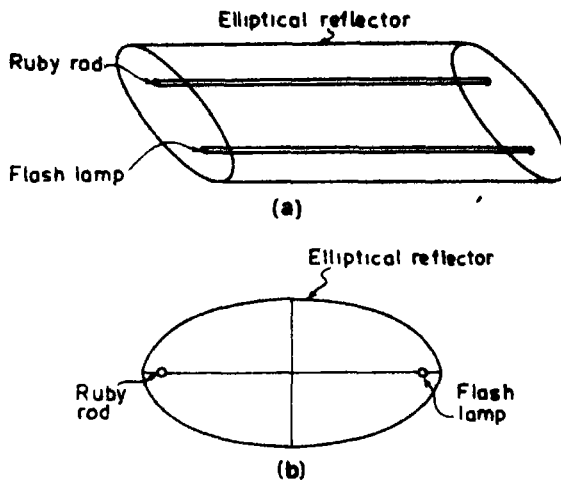



Fig 9 12 (a) The flashlamp and the ruby rod coincide with the focal lines of a cylindrical reflector with elliptical cross section This configuration for pumping leads to an efficient transfer of energy from the lamp to the ruby rod (b) The transverse cross section of the system

reflection from the reflector surface focuses to the other focus This helps in an efficient transfer of energy from the flashlamp to the laser rod

9 4 *The Helium–Neon Laser*

The helium–neon laser was the first gas laser to be operated successfully it was fabricated by Ali Javan and his coworkers in Bell Telephone Laboratories in the USA in 1961

The pumping mechanism used in solid state lasers is usually the optical pumping in which the lasing medium is pumped to a state of population inversion with the help of a flash lamp or a high-power continuous lamp Optical pumping methods would be very inefficient for gas lasers because gas atoms are characterized by sharp energy levels compared to solids, which are characterized by energy bands rather than energy levels† Thus for laser systems employing gases, one generally uses

† There is however an exception—namely, when the emission radiation of one gas atom corresponds exactly to the frequency required to excite another gas atom One laser has been made to operate in this way A discharge through helium results in emission at a wavelength of 388 8 nm, which matches almost exactly with the wavelength required to excite cesium atoms to an excited level Thus, the passage of the emission from helium discharge through an atmosphere of cesium atoms would raise the atoms to an excited level and thus lead to population inversion between the excited level and another lower lying energy level

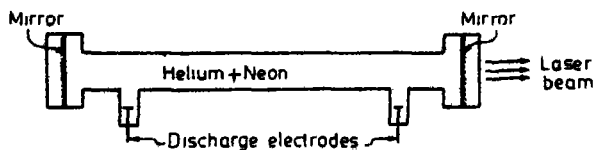



Fig 9 13 Schematic of a helium–neon laser

the electrical discharge as an efficient method of producing population inversion

The helium–neon laser consists of a long and narrow discharge tube (of diameter of about 1 cm and about 80 cm long) filled with about 1 torr† of helium and about 0 1 torr of neon (see Fig 9 13) The gas mixture of helium and neon forms the lasing medium and this mixture is enclosed between a set of mirrors forming a resonant cavity, one of the mirrors is completely reflecting and the other partially reflecting so as to couple out the laser beam

When a discharge is passed through the gaseous mixture, electrons are accelerated down the tube These accelerated electrons collide with the helium atoms and excite them to higher energy levels Figure 9 14 shows some of the energy levels of helium and neon The helium atoms are efficiently excited to the levels F_2 and F_3 These levels happen to be metastable and hence the helium atoms excited to these levels spend a

† Torr is a unit of pressure 1 torr = 1 mm of Hg

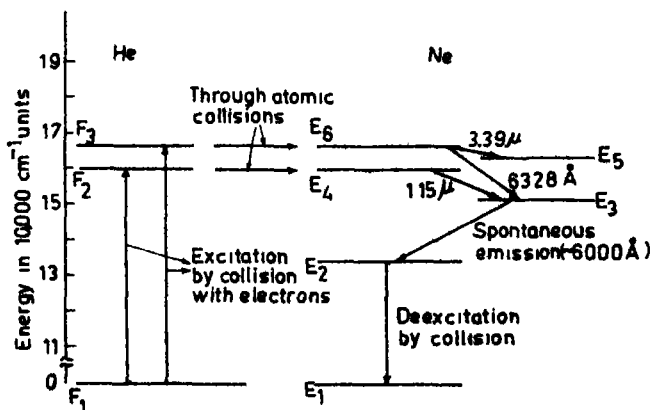



Fig 9 14 Energy level diagram showing the low-lying energy levels of helium and neon The excited helium atoms transfer their energy to the neon atoms through a collision process, thus creating a state of population inversion between levels E_4 (or E_6) and E_5 and E_3

sufficiently long amount of time before getting deexcited As shown in Fig 9 14 some of the excited states of neon correspond approximately to the same energy as the energy of the excited levels F_2 and F_3 of helium Thus, when helium atoms in levels F_2 and F_3 collide with neon atoms in the ground level E_1, an energy exchange takes place This results in the excitation of neon atoms to the levels E_4 and E_6 and deexcitation of the helium atoms to the ground level Because of the long lifetimes of the levels F_2 and F_3 of helium, this process of energy transfer has a high probability Thus, the discharge through the gas mixture continuously populates the neon excited energy levels E_4 and E_6 This helps to create a state of population inversion between the levels E_4 (or E_6) and the lower-lying energy levels E_5 and E_3 The various transitions (as shown in Fig 9 14) lead to emission at wavelengths of 3 39 μm, 1 15 μm, and 6328 Å, the first two correspond to the infrared region while the last wavelength corresponds to the well-known red light from the helium-neon laser Specific frequency selection may be obtained by employing mirrors which reflect only a small band of frequencies about the frequency of interest

The excited neon atoms drop down from the level E_3 to the level E_2 by spontaneously emitting a photon around $\lambda = 6000$ Å The pressures of the two gases in the mixture are so chosen that there is an efficient transfer of energy from the helium to the neon atoms Since the level marked E_2 is metastable, there is a finite probability of the excitation of neon atoms from E_2 to E_3 leading to a quenching of the population inversion When a narrow tube is used, the neon atoms in the level E_2 collide with the walls of the tube and get deexcited to the level E_1 Typical power outputs of helium-neon lasers lie between 1 and 50 mW of continuous wave for inputs of about 5–10 W

The light from gas lasers as compared to that from solid state lasers is found to be more directional and much more monochromatic† This is due to the various imperfections present in the solids and also the heating caused by the flash lamp The gas lasers are also capable of supplying a continuous laser beam without the need for elaborate cooling arrangements

One disadvantage of using internal mirrors (i e , with mirrors sealed inside the discharge tube) is that the mirrors are usually eroded by the gas discharge and have to be replaced Having the resonator mirrors external to the laser cavity also provides for greater flexibility However, when external mirrors are used, the ends of the discharge tube also cause a loss

† The light emerging from the helium-neon laser was so monochromatic that special techniques had to be evolved to measure such short linewidths

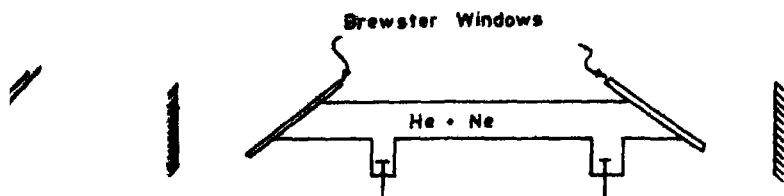



Fig 9 15 Helium neon laser with Brewster windows and external adjustable mirrors

due to reflection Thus one usually has two windows at the ends of the discharge tube which are at the Brewster angle (see Fig 9 15) For such windows, light which is polarized in the plane of the figure passes through without suffering any reflection loss while light with a perpendicular polarization suffers reflection losses When such windows are used, the output laser beam is polarized Also, since the Brewster windows produce asymmetry of the resonator, the modes are usually with x, y symmetry (see Fig 6 4) rather than with rotational symmetry

9 5 Four-Level Solid State Lasers

As we have already discussed, ruby laser is a three-level laser system and hence is not very efficient There exist various other kinds of solid state lasers, which are based on the four-level scheme (see Section 4 3) A very useful laser is one employing various energy levels of the neodymium ion Neodymium ions may be situated in various host materials like glass or yttrium aluminum garnet ($Y_3Al_5O_{12}$—abbreviated as YAG) etc The energy level diagram of neodymium ions is shown in Fig 9 16 Neodymium ions are excited to energy levels above that marked M with the help of a flashlamp or a continuous wave lamp These ions relax down to the level marked M which happens to be a metastable level with a lifetime of about 0 25 msecs Laser action is usually obtained between the levels M and L at a wavelength of about 1 06 μm in the infrared region

Typical threshold of a few hundred watts continuous wave power inputs are required for attainment of population inversion Output powers of several hundred watts have been obtained by using several laser rods in cascade

Neodymium-doped YAG and neodymium-doped glass lasers find widespread applications in industry Neodymium glass lasers can supply pulses of 100 psec duration and intensities as high as 10^{16} W/cm^2, such systems are being used to study laser-induced fusion reactions (see Chapter 11)

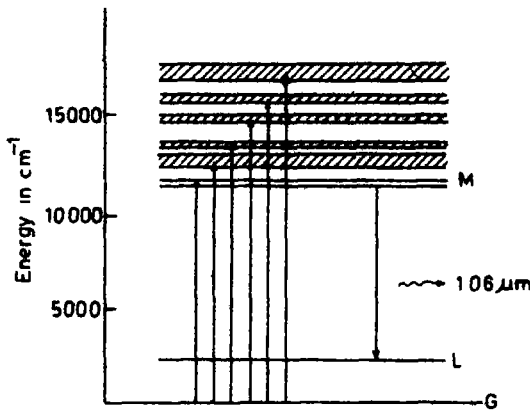



Fig 9 16 Energy level diagram of neodymium ions Laser emission occurs at a wavelength of about 1 06 μm

9 6 The Carbon Dioxide Laser

A molecule is made up of two or more atoms bound together Thus, in addition to the electronic motions, atoms in the molecule may vibrate in different modes or rotate about various axes For example, Fig 9 17 gives the three modes of vibration of a molecule of carbon dioxide The molecule is thus characterized not only by electronic levels but also by vibrational and rotational levels, each electronic level is split into various vibrational sublevels (due to the vibrational motions) and each vibrational level is further subdivided into rotational sublevels The energy difference between various electronic levels corresponds to the visible and the ultraviolet while the energy difference between various vibrational levels

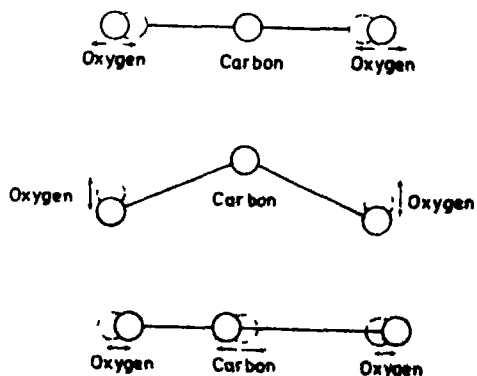



Fig 9 17 The three modes of vibration of a carbon dioxide molecule

corresponds to the infrared region and the energy difference between various rotational levels corresponds to the far infrared region of the spectrum

In the carbon dioxide laser, one makes use of transitions between a rotational sublevel of a vibrational level and a rotational sublevel of a lower vibrational level The population inversion is achieved through a gas discharge and the lasing action produces outputs at 10 6 μm and at 9 6 μm The addition of N_2 gas to the carbon dioxide gas is seen to increase the efficiency of the carbon dioxide laser Thus, the nitrogen molecules are excited to the upper vibrational level and it is seen that the energy difference between this level and the vibrational ground level of the nitrogen molecule matches very well with the energy required to excite the carbon dioxide molecule to the upper laser level Thus there is a very efficient transfer of energy between a nitrogen molecule and a carbon dioxide molecule, resulting in the excitation of the carbon dioxide molecule (see Fig 9 18) This increases the efficiency of the carbon dioxide laser

Carbon dioxide lasers are much more efficient as compared to other gas lasers This is because of the fact that in other gas lasers, the deexcitation from the lower laser level to the ground level involves a sufficient amount of energy which does not contribute to the output laser beam In contrast in the carbon dioxide laser, the levels taking part in

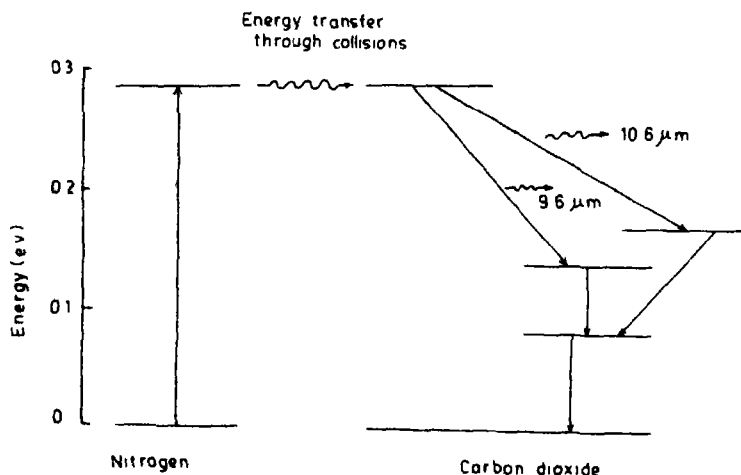



Fig 9 18 Energy levels of nitrogen and carbon dioxide Transfer of energy from nitrogen to carbon dioxide occurs through collisions resulting in excitation of carbon dioxide molecules leading to a population inversion The important lasing transitions occur at $\lambda \sim$ 9 6 and 10 6 μm

laser transitions are the vibrational rotational levels of the lowest electronic level These levels are very close to the ground level and hence a large portion of the input energy is converted into the output laser energy, resulting in very high efficiencies

Output powers of several watts to several hundred watts may be obtained from carbon dioxide lasers High-power carbon dioxide lasers should find applications in industry for welding, hole drilling, cutting, etc because of the very high output power

The wavelength at which CO_2 lasers lase falls in a band where atmospheric attenuation is very little Hence carbon dioxide lasers should also find applications in open air communications system (see Chapter 12) They must also be useful in optical radar systems

9 7 Dye Lasers

Dyes used in lasers are organic substances which absorb in the near ultraviolet, visible, or the near infrared regions of the electromagnetic spectrum These substances are most commonly solids which are dissolved in various solvents like water, ethyl alcohol, methanol, ethylene glycol, etc Great interest was generated in lasers using these dyes because of their broad absorption and fluorescent spectrum and hence the possibility of realizing a cheap and simple tunable laser In addition one can also choose from a large range of dyes for emission in any region of the visible spectrum

The levels taking part in the lasing transition are the different vibrational sublevels of different electronic states of the dye molecule For example, Fig 9 19 gives a typical energy level diagram of a dye molecule in which S_0 represents the electronic ground level which contains a large number of sublevels corresponding to the various vibrational and rotational states of the molecule The dense collection of the rotational sublevels leads to a near continuum of energy states Each electronic state is characterized by a similar broad continuum of states This is what leads to the characteristic broad absorption and emission spectrum of a dye molecule

Because of absorption of light, dye molecules get excited from the ground state S_0 to higher vibrational rotational levels of the next electronic state S_1 Because of thermal redistribution in the level S_1 (which takes place in times of about 10^{-11} sec), most of the dye molecules drop down to the lowest vibrational level B of S_1 Radiation is emitted when the molecules decay from the level B to any higher-lying vibrational sublevel of S_0 This is termed fluorescence The lifetime τ of level B for

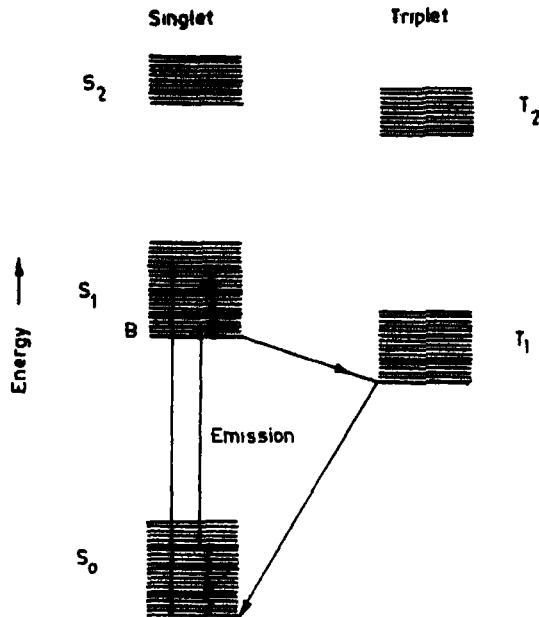



Fig 9 19 Typical energy level diagram of a dye molecule S_0 represents the lowest electronic level (a singlet) S_1, S_2, etc represent excited singlet levels, T_1, T_2, etc represent the triplet levels

dye molecules is about 10^{-9} sec (for solid state and other laser media, the lifetime of the upper level is about 10^{-5}–10^{-3} sec) As is evident, the peak wavelength of the emitted flourescence spectrum is higher than that of the absorption spectrum

Molecules from the state S_1 can also make a nonradiative relaxation to the triplet level T_1, this is referred to as intersystem crossing This transition process is detrimental to laser action for two reasons Firstly, this process leads to a reduction of the population of level S_1 which is the upper laser level Secondly, the absorption spectrum of $T_1 \rightarrow T_2$ usually overlaps the emission spectrum of $S_1 \rightarrow S_0$, thus leading to a loss at the wavelength corresponding to the laser emission In fact this loss can be so strong as to even prevent laser oscillation For good laser action, one must reach the threshold level before a significant number of molecules have dropped to level T_1 This requires an exciting source with a very fast rise time, which could either be flash lamps with very fast rise times or pulsed lasers

Tuning in dye lasers may be obtained using various techniques, one interesting method is to replace one of the mirrors of the cavity with a

diffraction grating which is adjusted so that the first-order reflected beam for the desired wavelength is along the axis of the optical cavity By rotating the grating, the wavelength can be altered In fact the grating behaves as a mirror with a sharp reflectance spectrum Using such techniques, tuning over 500 Å has been obtained

In Section 6 7 we had discussed the concept of mode locking Since the emission bands of dye lasers are usually very large (~100–1000 Å, i e , 10^{13}–10^{14} Hz), one can indeed obtain mode-locked picosecond pulses with the additional freedom of wavelength tuning

For further details on dye lasers, readers may look up Schafer (1973)

9.8. *Semiconductor Lasers*†

Semiconductors, as the name implies, are materials that have electrical conductivities lying somewhere between those of conductors and insulators In a solid, the atoms lie very close to each other and the presence of one atom has an effect on the other atoms As the atoms are brought together to form a solid, the effect is that the atomic energy levels get perturbed and one ends up with energy bands rather than energy levels (see Fig 9 20) In an energy band, the energy levels are so close together that it is essentially continuous

If we consider, for example, sodium, then the sodium atom consists of a nucleus surrounded by various orbiting electrons The outermost orbit has a single electron When the sodium atoms are brought close

† A more detailed account on semiconductor lasers has been given in the Nobel lecture of Basov which is reprinted in Part III of this book

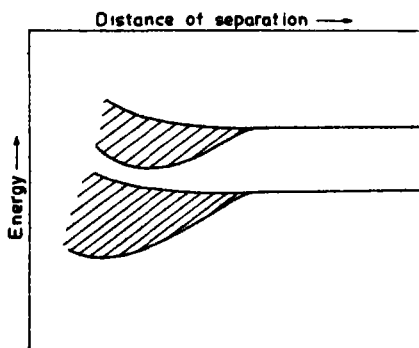



Fig 9 20 Atoms are characterized by discrete energy levels When they are brought together to form a solid, then the energy levels get perturbed and one finally obtains energy bands rather than energy levels

together to form the sodium metal, these outer electrons detach themselves from the atom and are free to move in the metal The fact that a solid piece of sodium has a large number of free electrons moving in the crystal leads to the metallic properties of sodium Thus when an electric field is applied across sodium, the electrons flow and constitute a current

In contrast, in certain other materials, called insulators, no such free electrons exist in the solid and they do not conduct electricity

In between conductors and insulators, lie a very interesting and important class of materials called semiconductors In these materials, when the temperature is extremely low, there are no free electrons in the solid and they behave as insulators On increasing the temperature, some electrons gain enough energy to get detached from the atom As the temperature is increased, more and more electrons are able to free themselves and thus the conductivity of these materials increases with the increase of temperature This is in complete contrast with conductors, whose resistance increases with increase of temperature

Typical examples of the class of semiconductors are germanium and silicon The outermost orbits of these atoms contain four electrons When a large number of such atoms are brought together to form a solid, the adjacent atoms share one electron each and form between them what are called bonds (see Fig 9 21) When the temperature of the solid increases, electrons tend to break up from the bond and become free This leads to a vacant site at which the electron was previously residing These vacant sites are called holes and behave just like electrons but with a positive charge When a voltage is applied across such a material, the electrons tend to flow to the positive terminal and the holes tend to flow towards the negative terminal Their flow leads to a flow of current

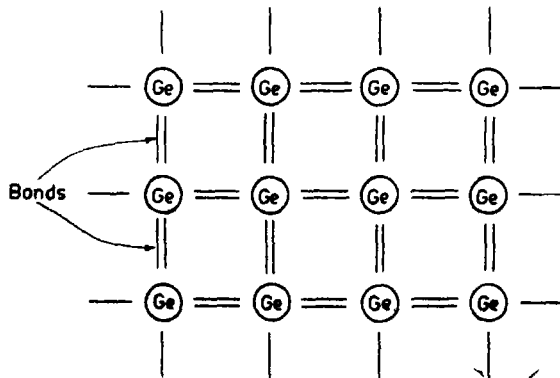
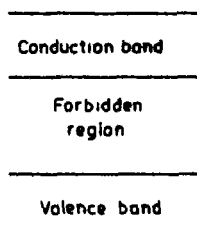



Fig 9 21 Germanium atoms are held together by covalent bonds formed by sharing of electrons between two germanium atoms

Conduction band

Forbidden region

Valence band

Fig. 9.22 Band structure in a semiconductor. At 0°K, the valence band is completely filled while the conduction band is completely empty. As the temperature is increased, electrons from the valence band may gain enough energy and make a transition to the conduction band.

The energy bands of a semiconductor are separated by regions of energy in which no energy levels exist. This region is called the forbidden region and the energy width of this band is called the band gap. Typical band gap energies are 0.76 eV for germanium and 1.1 eV for silicon.† At absolute zero, the electrons occupy the lowest energy levels, and for semiconductors, this leads to a completely filled band separated by a completely empty band. The filled band represents the valence band and the next upper empty band is termed the conduction band. When the temperature of the semiconductor is raised from absolute zero, some of the electrons at the top of the valence band may gain enough energy to make a transition to the conduction band. This leaves behind a hole in the valence band. The excited electron and the hole can take part in conduction. If by some process we are able to excite a large number of electrons to the conduction band, then within a very short time the electrons in the valence band and the conduction band may rearrange and result in the creation of a large number of holes in the top of the valence band and a large number of electrons in the bottom of the conduction band (see Fig. 9.23). Then, if a photon having energy slightly greater than the band gap energy is incident (or produced spontaneously), such a photon will stimulate a large number of downward transistions of the electrons from the conduction to the valence band. This would result in a coherent amplification (see Fig. 9.23). If the resulting medium is enclosed between two mirrors, one may have a laser oscillator.

An easy method to obtain population inversion is through the use of p–n junctions. We will now briefly discuss the p-type and the n-type semiconductors and how they may be used in obtaining population inversion.

In the earlier discussion we had considered a pure semiconductor. In such a semiconductor one has an equal number of electrons and holes

† At room temperature ($T \simeq 300$°K), $kT \simeq 0.025$ eV and since the band gap energies are comparable to kT we may have considerable excitation of electrons from the valence to the conduction band. For an insulator the band gap energies are considerably larger; for example, for diamond this energy is about 6 eV.

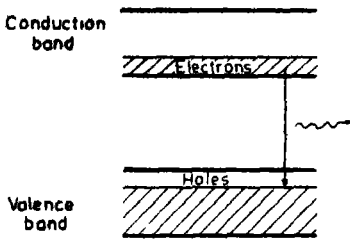



Fig 9 23 An electron and hole may combine and in the process emit a photon

Consider the effect of replacing some of the atoms (of, say, germanium) by atoms of another element such as arsenic possessing five electrons in the outermost orbit Since only four electrons are required for the formation of bonds, the fifth electron is not tightly bound and is free to move in the crystal Thus the addition of impurity atoms containing five outermost electrons results in a crystal possessing an excess of electrons Such a semiconductor is termed an n-type semiconductor Similarly, if we add impurity atoms of, say, gallium, which has only three outermost electrons, then one creates an excess of holes in the crystal Such a semiconductor is termed a p-type semiconductor

If we now form a junction between a p-type and an n-type semiconductor (see Fig 9 24) then the excess electrons in the n-type region tend to flow across the junction into the p-type region and the excess holes in the p region tends to flow to the n region This creates an electric field opposing this flow and the flow stops

If we now connect the p-type semiconductor to the positive terminal of a direct current source and the n-type to the negative terminal, then one produces a flow of electrons (from the n region) and holes (from the p region) into the junction region These electrons and holes may recombine to produce radiation This amounts to the deexcitation of an electron from the conduction band to the valence band If the current is large enough one may have laser action A resonant cavity is formed by just cleaving the junction ends Because the refractive index of a semiconductor is quite large ($\simeq 3\ 6$), the reflectivity at the crystal–air interface is as high as 32% One may, of course, coat the surfaces for reducing the threshold current

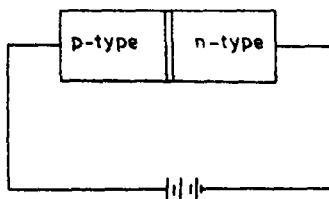



Fig 9 24 A p–n junction

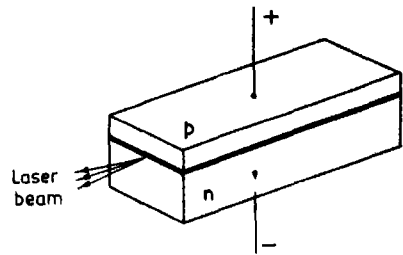



Fig 9 25 A p–n junction laser The laser emission is confined to a very narrow region around the junction

The laser emission is restricted to a very thin region ($\simeq 1\ \mu$m) around the junction (see Fig 9 25) Thus, the laser beam usually emerges over a wide angle Also, in comparison to the gas laser, the laser emission occurs between two bands of energies (as compared to two well-defined energy levels in gas lasers) and hence the emission is not as monochromatic as that from a gas laser

The semiconductor laser is extremely small in physical dimension Because of the direct conversion of electrical current into light energy, it is also very efficient It also lends itself to direct modulation schemes Thus by modulating the diode current, the output beam may be modulated

Because of these special characteristics of semiconductor lasers, they are expected to find widespread applications Thus, they seem to hold great promise as sources for light wave communication systems (see Chapter 12) The emission wavelength of some of the semiconductor lasers (like GaAs and GaInAsP) match with the wavelength region where fibers show very little attenuation They are also expected to find applications in portable optical radar equipment space communications etc

In the above, we have discussed the basic principle behind p–n junction lasers These lasers can be operated continuously only at very low temperatures (typically at liquid nitrogen temperature $\simeq 77°$K) Their operation at room temperature may be obtained only in a pulsed form Recently, this problem has been overcome by the construction of what are called heterojunction lasers Such lasers may be operated continuously even at room temperatures with low threshold current densities Such an operation has been achieved by the combined effect of a well-confined active region where a population inversion exists and by confining the oscillating laser beam The heat produced is also dissipated away efficiently by cementing the diode to a tin plate which acts as a heat sink For details on semiconductor lasers, the reader is referred to Kressel (1979) and Kressel and Butler (1977)

PART II

Some Important Applications of Lasers

10

Spatial Frequency Filtering and Holography

10.1 Introduction

One of the most interesting and exciting application of lasers lies in the fields of spatial frequency filtering and holography. In this chapter, we briefly outline the principle behind spatial frequency filtering and holography and briefly discuss their applications. A nice account on holography can also be found in the Nobel lecture by Gabor, which is reproduced at the end of the book.

10.2 Spatial Frequency Filtering

Just as the Fourier transform of a time-varying signal gives its temporal frequency spectrum, similarly the spatial Fourier transform of a spatially varying function (like the transmittance of an object) gives the spatial frequency spectrum of the function (see Appendix A). It can indeed be shown that the field distribution produced at the back focal plane of an aberrationless converging lens is the two-dimensional Fourier transform of the field distribution in the front focal plane of the lens (see, e.g., Ghatak and Thyagarajan, 1978). Thus, if $f(x, y)$ represents the object distribution in the front focal plane of an aberrationless converging lens, then the field distribution in the back focal plane is given by (see Appendix E)

$$g(x, y) = \frac{i}{\lambda f} \iint f(x', y') \exp\left[\frac{2\pi i}{\lambda f}(xx' + yy')\right] dx'\, dy' = \frac{i}{\lambda f} F\left(\frac{x}{\lambda f}, \frac{y}{\lambda f}\right) \qquad (10.2\text{-}1)$$

where $F(x/\lambda f, y/\lambda f)$ represents the Fourier transform of $f(x, y)$ evaluated at the spatial frequencies $(x/\lambda f, y/\lambda f)$—see Appendix A, f represents the focal length of the lens and λ the wavelength of illumination Thus if on the front focal plane is placed an object with a transmittance of the form

$$f(x, y) = A\cos\left(\frac{2\pi x}{a}\right) = \frac{A}{2}\left[\exp\left(\frac{2\pi i x}{a}\right) + \exp\left(-\frac{2\pi i x}{a}\right)\right] \quad (10\ 2\text{-}2)$$

then on the back focal plane one would obtain a field distribution given by

$$g(x, y) = \frac{Ai}{2\lambda f}\left[\delta\left(\frac{x}{\lambda f} + \frac{1}{a}\right) + \delta\left(\frac{x}{\lambda f} - \frac{1}{a}\right)\right]\delta(y) \quad (10\ 2\text{-}3)$$

where we have used the fact that the Fourier transform of $\exp(2\pi i x/a)$ is $\delta(u + 1/a)$, where u is the spatial frequency, which in the present case is $x/\lambda f$ Equation (10 2 3) represents the field corresponding to two bright dots at $(x = \lambda f/a,\ y = 0)$ and $(x = -\lambda f/a,\ y = 0)$

Now, the Fourier transform of the Fourier transform of a function is the original function itself except for an inversion, i e ,

$$\mathscr{F}[\mathscr{F}\{f(x)\}] = f(-x) \quad (10\ 2\text{-}4)$$

where the symbol $\mathscr{F}[\ ]$ stands for the Fourier transform of [ ]—see Appendix A Thus if another converging lens L_2 is placed such that the back focal plane of the first lens L_1 is the front focal plane of the second lens (see Fig 10 1), then on the back focal plane of L_2 one would obtain the original object distribution except for an inversion The resultant field distribution on the plane P_3 can be controlled by suitably placing apertures on the back focal plane P_2 and thus performing operations on the spatial frequency spectrum of the object The various apertures, stops, etc , that are placed in the plane P_2 are referred to as filters Thus, for example, a low-pass filter would be one which allows low spatial frequencies to pass through while blocking the high spatial frequencies This could, for example, be a screen with a small hole at the axis of the system Similarly, a high-pass filter would transmit all high spatial frequencies while blocking the low frequencies One could also have complex filters which alter both the amplitude and phase of the various spatial frequency components of the image

As an example, let us consider an object with an amplitude variation of the form

$$f(x) = A\cos(2\pi\alpha x) + B\cos(2\pi\beta x) \quad (10\ 2\text{-}5)$$

The object represented by Eq (10 2-5) has two spatial frequencies α and β If such an object is placed in the plane P_1 and illuminated by a

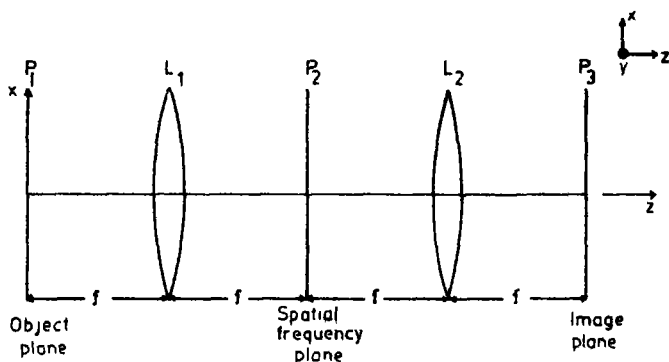



Fig 10 1 When an object transparency is placed in the front focal plane P_1 of an aberrationless converging lens L_1 and illuminated by a parallel beam of light, then on the back focal plane P_2 one obtains a spectrum of the spatial frequency components present in the object If a second lens L_2 is placed such that the plane P_2 coincides with its front focal plane then in the back focal plane P_3 of L_2 one obtains the image pattern corresponding to the spatial frequency spectrum in the plane P_2 One can indeed control the spatial frequency spectrum that is responsible for forming the image on the plane P_3 by placing filters on the plane P_2

coherent beam of light, then in the plane P_2, one would obtain four spots at distances $x = +\lambda f\alpha$, $+\lambda f\beta$, $-\lambda f\alpha$, and $-\lambda f\beta$ as shown in Fig 10 2 If we do not place any obstructions on the plane P_2, then both the spatial frequencies contribute in forming the image in the plane P_3 and one obtains in the plane P_3 the same amplitude distribution as that in the plane P_1 Now, consider placing two stops at the points $x = +\lambda f\alpha$ and $x = -\lambda f\alpha$ on the y axis in the plane P_2 Thus, no light from these points is allowed to reach the lens L_2 The lens L_2 sees only the spatial frequency β Hence it follows that the image pattern in the plane P_3 will be proportional to $\cos(2\pi\beta x)$ Thus, by placing stops in the plane P_2, we have been able to filter out the frequency component α This is the basic principle behind spatial frequency filtering As a corollary, we may mention that if we put a stop on the axis then it will filter out the low-frequency components This can also be seen from the fact that if a plane wave propagating parallel to the axis (which is associated with zero spatial frequency) falls on a lens, it gets focused to the axis on the plane P_2 and if we put a small stop on the axis in the plane P_2, then there will be no light reaching the lens L_2

Spatial frequency filtering finds widespread applications in various fields, we will discuss briefly some of these applications

If one looks closely at a newspaper photograph, one can immediately see that the image is in fact made up of a large number of closely

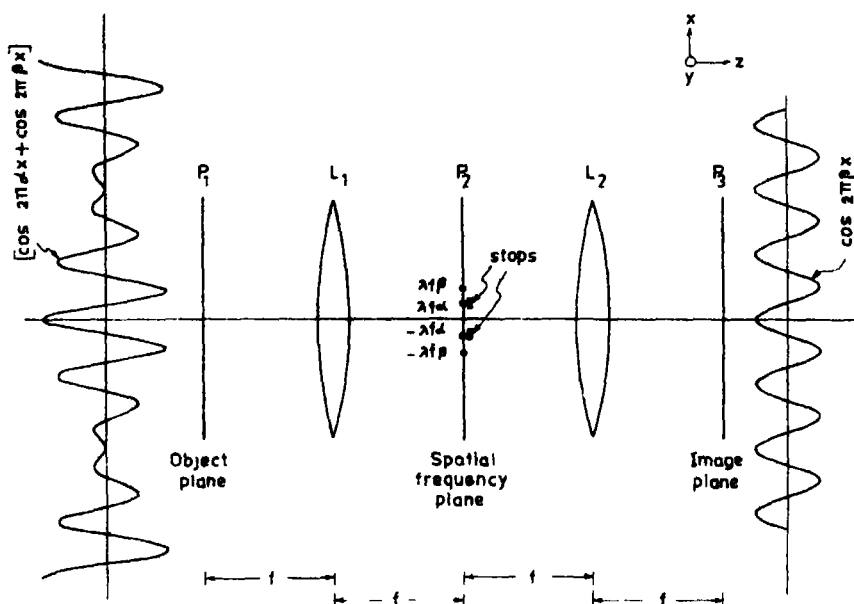



Fig 10 2 If we place an object with a transmittance proportional to Eq (10 2-5) on the plane P_1 then in the plane P_2 we will obtain four spots (on the y axis) at $x = \pm\lambda f\alpha$ and $\pm\lambda f\beta$ If we place two stops behind the spots at $x = \pm\lambda f\alpha$ then in the plane P_3 we would obtain an image pattern represented by $\cos(2\pi\beta x)$ Thus the frequency component α has been filtered out

arranged dots These closely arranged dots represent a high spatial frequency while the general image formed by these dots represents low-frequency components Thus these dots can be got rid of by spatial frequency filtering For example Fig 10 3a shows a photograph which consists of regularly spaced black and white squares The spatial frequency spectrum of the object is shown in Fig 10 3b If we place a screen with a small hole at the center on the back focal plane then the image produced is devoid of these dot patterns (see Fig 10 3c) By placing a small hole on the axis, one has essentially filtered out the high-frequency components in the object

Another application of spatial frequency filtering is in contrast enhancement When there is a large amount of background light in an image, the contrast in the image is poor Since the background light represents a distribution of zero spatial frequency, if we place the object in the front focal plane and in the back focal plane put a small stop on the axis (which cuts off the low-frequency components), then since the stop removes the low-frequency components one would obtain on the back

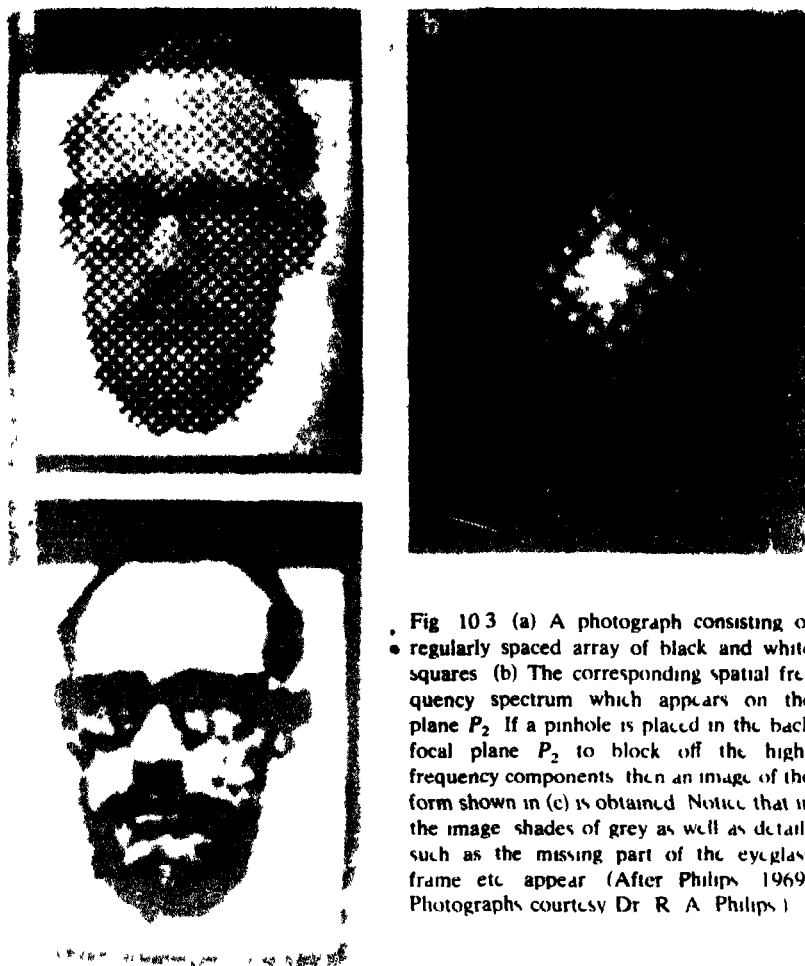



Fig 10 3 (a) A photograph consisting of regularly spaced array of black and white squares (b) The corresponding spatial frequency spectrum which appears on the plane P_2 If a pinhole is placed in the back focal plane P_2 to block off the high frequency components then an image of the form shown in (c) is obtained Notice that in the image shades of grey as well as details such as the missing part of the eyeglass frame etc appear (After Philips 1969 Photographs courtesy Dr R A Philips)

focal plane of L_2 an image with a much better contrast Such a process is termed contrast enhancement

Spatial frequency filtering can also be used for detecting nonperiodic (i e random) errors in a periodic structure Thus one could either transmit all the spatial frequencies corresponding to the periodic array and stop most of the light from the random noise or one could block the light corresponding to the periodic array frequencies and transmit most of the light corresponding to the defects Such techniques have indeed been used in photo mask inspection, electron tube grid inspection, etc (see e g Gagliano *et al*, 1969)

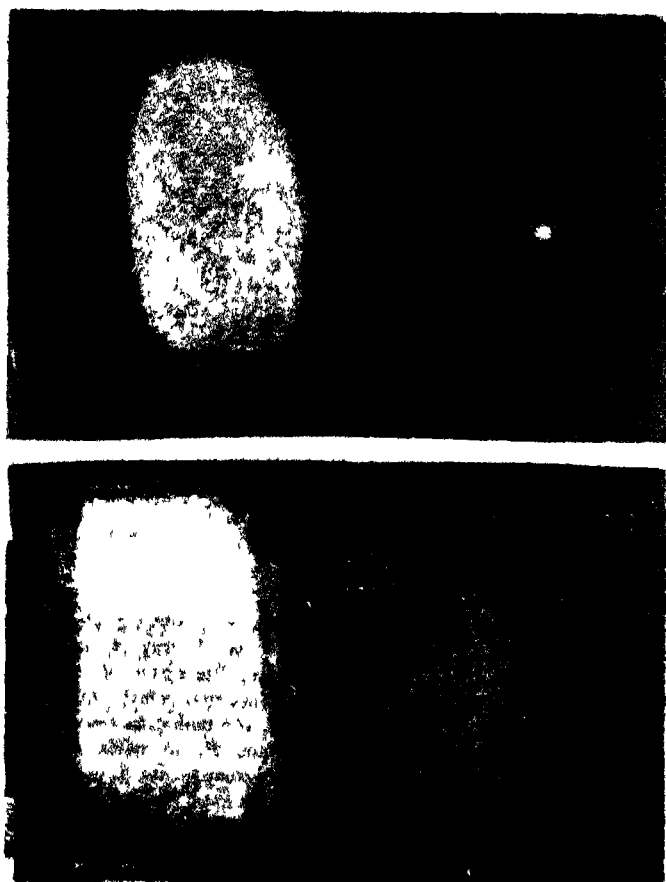



Fig 10 3 Fingerprint identification using optical cross correlation In the upper part the two fingerprints are matched, which results in appearance of a bright spot of light In the lower part the two fingerprints do not match and the resulting image is a smear (After Tsujiuchi *et al* 1971)

Another important application of spatial frequency filtering is in character recognition problems where it is necessary to detect the presence of certain characters in an optical image Here the filter is a complex filter produced using holographic principles and the output from the optical system is such that corresponding to the positions where the object contains the desired character, one obtains bright spots of light in the image† As an example we show in Fig 10 4 the identification of a

† For a detailed theoretical analysis of the character recognition problem see e g , Ghatak and Thyagarajan (1978)

fingerprint by this technique. In the top portion, the fingerprints are matched and one obtains a bright spot of light, when the two fingerprints do not match (lower portion), there is no appearance of a bright spot but only a smear. Character recognition problems will also find application in military defense, where it might be necessary to identify certain objects of interest. For further details on optical data processing, one may look up Cassent (1978).

10.3. Holography

An ordinary photograph represents a two-dimensional recording of a three-dimensional scene. The emulsion on the photographic plate is sensitive only to the intensity variations, and hence while a photograph is recorded the phase distribution which prevailed at the plane of the photograph is lost. Since only the intensity pattern has been recorded, the three-dimensional character (e.g., parallax) of the object scene is lost.

It was in the year 1948 that Dennis Gabor conceived of an entirely new idea and proposed a method of recording not only the amplitude but also the phase of the wave. The principle behind the method is the following. During the recording process, one superimposes on the wave (emanating from the object) another coherent wave called the reference wave. The two waves interfere in the plane of the recording medium and produce interference fringes. This is known as the recording process. The interference fringes are characteristic of the object and the recording medium records the intensity distribution in the interference pattern. This interference pattern has recorded in it not only the amplitude distribution but also the phase of the object wave. Thus, let

$$O(x, y) = O_0(x, y)e^{i\varphi(x,y)} \tag{10.3-1}$$

represent the field produced due to the object wave at the plane of the recording medium, $O_0(x, y)$ is the amplitude part and $\varphi(x, y)$ the phase part. Similarly, let

$$R(x, y) = Ae^{i\psi(x,y)} \tag{10.3-2}$$

represent the field produced due to the reference wave at the recording medium. Usually the reference wave is an obliquely incident plane wave, in which case A is a constant. The total field produced at the recording medium is

$$U(x, y) = O_0(x, y)e^{i\varphi(x,y)} + Ae^{i\psi(x,y)} \tag{10.3-3}$$

and the intensity pattern recorded by the recording medium would be

$$I(x, y) = |U(x, y)|^2 = O_0^2(x, y) + A^2 + O_0(x, y)A \exp\{i[\varphi(x, y) - \psi]\}$$
$$+ O_0(x, y)A \exp\{-i[\varphi(x, y) - \psi(x, y)]\} \qquad (10.3\text{-}4)$$

where we have omitted a constant of proportionality and have carried out a time averaging.† It can immediately be seen from the above that the recorded intensity distribution has the phase of the object wave $\varphi(x, y)$ embedded in it. Since the recorded intensity pattern has both the amplitude and phase recorded in it, Gabor called the recording a hologram (*holos* in Greek means "whole").

The hologram has little resemblance to the object. It has in it a coded form of the object wave. The technique by which one reproduces the image is termed reconstruction. In the reconstruction process, the hologram is illuminated by a wave called the reconstruction wave; this reconstruction wave in most cases is similar to the reference wave used for recording the hologram. When the hologram is illuminated by the reconstruction wave there emerges from the hologram various wave components, one of which is the object wave itself. In order to show this, we see that when the exposed recording medium is developed, then one, in general, gets a transparency with a certain transmittance. Under proper conditions, the amplitude transmittance of the hologram can be made to be linearly proportional to $I(x, y)$. Thus apart from some constants, one can write for the amplitude transmittance of the hologram

$$t(x, y) = I(x, y) \qquad (10.3\text{-}5)$$

If this transparency is illuminated with the reconstruction wave, the emerging wave would be given by

$$t(x, y)A \exp[i\psi(x, y)] = [O_0^2(x, y) + A^2]A \exp[i\psi(x, y)]$$
$$+ O_0(x, y)A^2 \exp[i\varphi(x, y)]$$
$$+ A^2 O_0(x, y) \exp\{-i[\varphi(x, y) - 2\psi(x, y)]\}$$
$$(10.3\text{-}6)$$

The second term indeed represents the original object wave apart from the constant multiplicative factor A^2. The first term represents the reconstruction wave itself but which has been modulated in amplitude

† We are assuming the fields to be monochromatic with a time dependence of the form $e^{i\omega t}$. Now for two functions f and g with time variations of the form $e^{i\omega t}$

$$\langle \text{Re} f \, \text{Re} g \rangle = \tfrac{1}{2} \langle \text{Re} f^* g \rangle$$

where angular brackets denote time averaging and Re f stands for the real part of the function f.

The last term represents the complex conjugate of the object wave The three wave components can be spatially separated by a proper choice of the reference wave

The second term, which represents a reproduction of the object wave, is identical to the wave that was emanating from the object when its hologram was being recorded Thus, when one views this wave (emerging from the hologram), then one sees a reconstructed image of the object in its true three-dimensional form Thus, as with the original object, one can move one's viewing position and look around the object If the hologram has recorded in it sufficient depth of field, one has to refocus one's eyes to be able to see distinctly the objects which are far away One can even place a lens on the path of the reconstructed wave and form an image of the object on a screen

In addition to the virtual image, the reconstruction process generates another image, which is a real image, this is represented by the third term in Eq (10 3-6) This real image can indeed be photographed by placing a suitable light-sensitive medium (like a photographic plate) at the position where the real image is formed

Although the principle of holography was laid down by Gabor in 1948, it was not until the lasers arrived in 1960 that holography attained practical importance Before the advent of the laser, one had to employ the method of in-line holography (as proposed by Gabor), in which the reference beam is approximately parallel to the object wave and the paths traversed by both the object wave and the reference wave are almost equal (see Fig 10 7), this was required because the existing sources like mercury discharge lamps had only small coherence lengths† The in-line technique has associated with it the disadvantage that the waves that form the virtual and real images travel along the same direction Thus, while viewing the virtual image, one is faced with an unfocused real image and conversely The early work on holography was in fact on the removal of this problem associated with this geometry of recording

It was in the year 1962 that Leith and Upatneiks introduced the technique of "off-axis holography," which overcame the difficulty associated with the in-line technique In the technique proposed by Leith and Upatneiks, one uses a reference beam which falls obliquely on the photographic plate (see Figs 10 5, 10 6, and Fig 8 of the Nobel lecture

† The high-pressure mercury arc lamp emits a green line at 5461 Å The coherence length of this line is, in fact only about 10 μm (The width of the line at 5461 Å is about 5×10^{12} Hz) On the other hand, the 6058-Å line emitted by krypton has a coherence length of ~ 20 cm, but the power output per unit area of this source is very low When one tries to increase the source area one loses spatial coherence The notion of coherence length has been discussed in Chapter 9

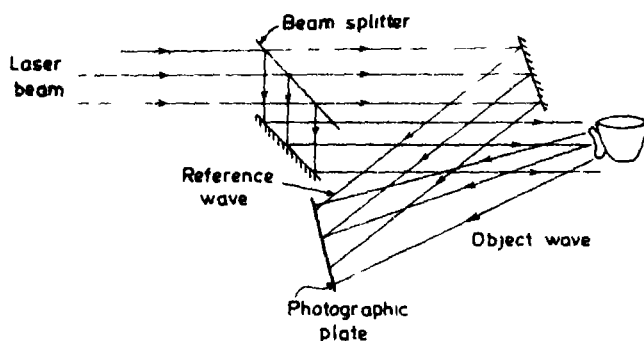



Fig 10 5 The figure shows an arrangement for the recording of a hologram The beam from a laser is split up into two portions one part is used to illuminate the object and the other part is used as a reference beam The waves scattered from the object interfere with the reference wave to form the hologram

by Gabor at the end of the book) Using such a technique in the reconstruction process, one obtains well-separated virtual and real images The use of such a technique was made possible by the large coherence length of the laser† The importance of coherence can be seen

† For further discussion on off-axis holography the reader is referred to Collier *et al* (1971) Ghatak and Thyagarajan (1978), and the Nobel lecture by Gabor which is reproduced at the end of the book

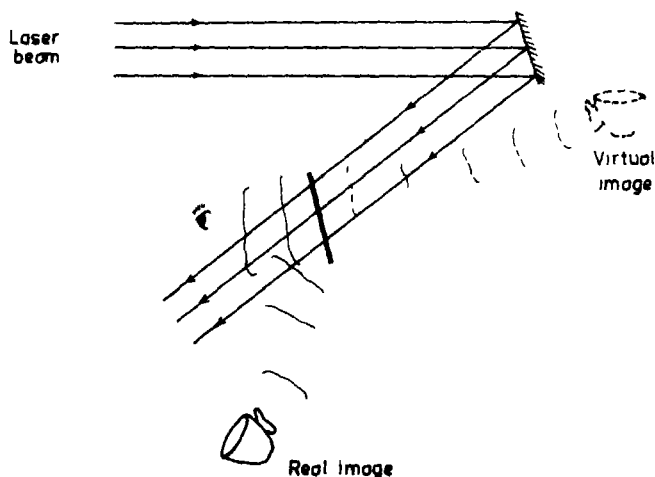



Fig 10 6 In the reconstruction process, the hologram is illuminated by a reconstruction wave, which in most cases is identical to the reference wave used for forming the hologram The reconstruction wave after passing through the hologram produces a real and a virtual image The virtual image can be viewed and exhibits all the true three-dimensional characteristics like parallax depth etc

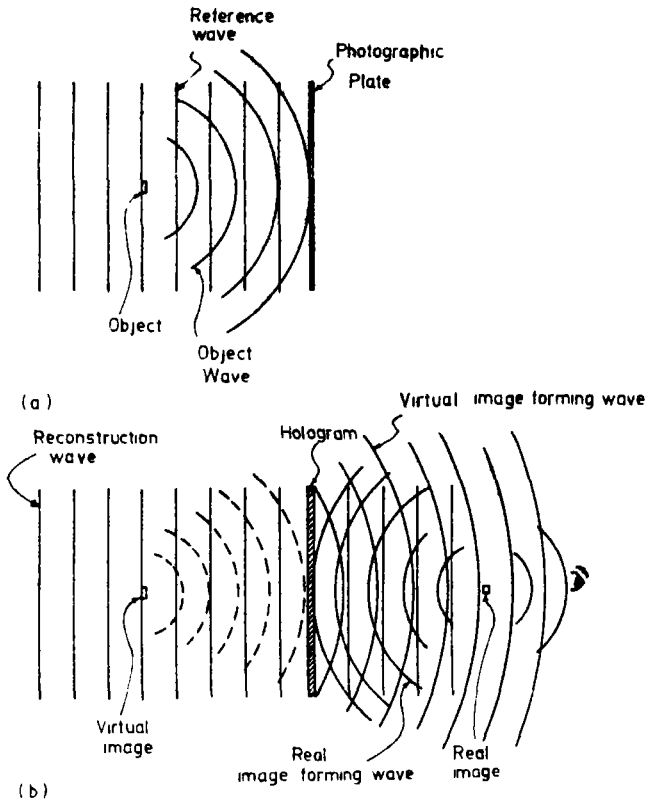



Fig 10 7 The in-line holography technique in which the object wave and the reference wave are traveling almost parallel to each other During reconstruction, both the waves producing the virtual and real images are traveling approximately in the same direction, this produces some difficulties while viewing the images

from the fact that holography is essentially an interference phenomenon Thus it is essential that the illuminating wave possess sufficient spatial coherence so that the wave from every object point may interfere with the reference wave As we had noted in Section 9 2, we can obtain a wave with sufficient spatial coherence by making use of pinholes for illuminating Alternatively, one could move the source far enough from the scene But both the above methods essentially decrease the available power The arrival of lasers overcame this difficulty Further, in order that stable interference fringes be formed in the hologram of the complete object scene to be recorded, the maximum path difference between the reference wave and the wave from the object must be less than the coherence length

Holograms exhibit very interesting properties For example when one records the hologram of a diffusely reflecting object, each point on the object scatters light on the complete surface of the hologram Thus, every part of the hologram receives light from all parts of the object Hence even if one breaks the recorded hologram into various parts, each part is capable of reconstructing the entire object, the resolution in the image decreases as the size of the hologram decreases In fact, when one records holograms of transparencies, one often uses a ground glass screen which enables the hologram to receive light from all parts of the transparency

The principle of holography finds applications in many diverse fields We will discuss a few of them here Numerous other applications can be found in Gabor's Nobel lecture

We had observed that information about depth can also be stored in a hologram Consider the problem of locating a transient event concerning a microscopic particle in a certain volume and studying it If one uses an ordinary microscope, then since the event is transient, it is, in general, difficult to first locate the particle and study it On the other hand, if one makes a holographic record of the complete volume then one can "freeze" the event in the hologram On reconstruction, there would emerge from the hologram the same wave with the difference that it is now not transient If one now uses a microscope one can easily locate the particle and study it at leisure

One of the most important applications of holography has been in interferometry In one technique called double exposure holographic interferometry, one first partially exposes an object to the photographic plate with a reference wave Now, the object is stressed, and one makes another exposure along with the same reference wave If the resulting hologram is developed and illuminated by a reconstruction wave then there would emerge from the hologram two object waves, one corresponding to the unstressed object and the other corresponding to the stressed object These two object waves would interfere to produce interference fringes Thus, on viewing through the hologram, one finds a reconstruction of the object superimposed with fringes The shape and number of fringes give one the distribution of strain in the object One can employ the above technique in nondestructive testing of objects Many experimental results and various other applications are discussed in the Nobel lecture by Gabor, which is reproduced at the end of the book

11

Laser-Induced Fusion

11 1. Introduction

It is well known that the enormous energy released from the sun and the stars is due to thermonuclear fusion reactions, and scientists have been working for over 40 years now to devise methods to generate fusion energy in a controlled manner Once it is achieved one will have an almost inexhaustible supply of relatively pollution-free energy A thermonuclear reactor based on laser-induced fusion offers great promise for the future With the tremendous effort being expended on fabrication of extremely high-power lasers, the goal appears to be not too far away, and once it is practically achieved it would lead to the most important application of the laser

In the next section, we discuss the basic physics behind the energy released in a fusion reaction, in Section 11 3 we discuss the laser energy requirements, and in Section 11 4 we briefly describe the laser-induced fusion reactor and some of the practical difficulties

11 2 The Fusion Process

A nucleus of an atom consists of protons and neutrons which are nearly of the same mass The proton has a positive electrical charge and the neutron, as the name implies, is electrically neutral Because neutrons and protons are the essential constituents of atomic nuclei, neutrons and protons are usually referred to by the general name "nucleon " If one assumes that the forces between the nucleons are of the Coulomb type, then the nucleons would have flown apart because of the repulsion between two protons and also because no Coulomb-type forces exist between two neutrons and between a neutron and a proton Since the nucleons are held together in the nucleus, there must be a short-range attractive force between them Indeed, it is believed that at short dis-

tances ($\leqslant 10^{-13}$ cm), very strong attractive forces exist between the nucleons† and this force is independent of the charge of the nucleons, i e , the force between two protons or two neutrons or between a proton and a neutron are essentially similar Because of the strong attractive forces between the nucleons, a certain amount of energy has to be supplied to split a nucleus into its constituent nucleons This is known as the binding energy‡ of the nucleus

Consider a nuclear reaction in which the two deuterons react to form a tritium nucleus and a proton

$$\mathrm{D} + \mathrm{D} \rightarrow \mathrm{T}(1\;01\,\mathrm{MeV}) + \mathrm{H}(3\;01\,\mathrm{MeV}) \qquad (11\;2\text{-}1)$$

The binding energy of each of the deuterium nuclei is 2 23 MeV and the total binding energy of the tritium nucleus is 8 48 MeV Thus, there is a net gain in the binding energy which is $8\;48 - 2 \times 2\;23 = 4\;02\,\mathrm{MeV}$ Physically, a loosely bound system goes over to a tightly bound system resulting in the liberation of energy, this energy appears in the form of kinetic energy of tritium and proton, which are given in parentheses in Eq (11 2-1) Nuclear reactions such as that represented by Eq (11 2-1) in which two loosely bound light nuclei produce a heavier tightly bound nucleus are known as fusion reactions§

† Beyond the range of this short-range force the forces are of Coulomb type

‡ The binding energy is calculated using the famous Einstein mass–energy relation

$$E = mc^2$$

where c ($\simeq 3 \times 10^{10}$ cm/sec) represents the speed of light in free space If Z and N represent the number of protons and of neutrons inside the nucleus, then the total binding energy, Δ, will be given by

$$\Delta = (Zm_p + Nm_n - M_A)c^2$$

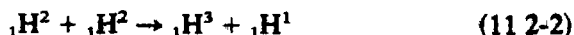
where m_n m_p and M_A represent the masses of the neutron the proton, and the atomic nucleus respectively For example the nucleus of the deuterium atom (which is known as the deuteron) has a mass of 2 01356 amu (1 amu $\simeq$ 1 661 $\times 10^{-24}$ g, which is equivalent to 931 5 MeV) Since deuteron consists of one proton and one neutron, one obtains

$$\begin{aligned}\Delta &- (1\;00728 + 1\;00866 - 2\;01356) \times 1\;661 \times 10^{-24} \times (3 \times 10^{10})^2\,\mathrm{erg}\\ &= 3\;56 \times 10^{-6} \times (1\;6 \times 10^{-12})^{-1} \times 10^{-6}\,\mathrm{MeV}\\ &\simeq 2\;23\,\mathrm{MeV}\end{aligned}$$

which represents the binding energy of the deuteron In the above equation, we have used 1 00728 amu and 1 00866 amu to represent the masses of proton and neutron, respectively

§ On the other hand, in a fission process a loosely bound heavy nucleus splits into two tightly bound lighter nuclei, again resulting in the liberation of energy For example, when a neutron is absorbed by a $_{92}\mathrm{U}^{235}$ nucleus, the $_{92}\mathrm{U}^{236}$ nucleus is formed in an excited state (the excitation energy is supplied by the binding energy of the absorbed neutron) This $_{92}U^{236}$ *nucleus may undergo fission to form nuclei of intermediate mass numbers* (like $_{56}\mathrm{Ba}^{140}$ and $_{36}\mathrm{Kr}^{93}$ along with three neutrons) The energy released in a typical fission reaction is about 200 MeV

Since deuterium and tritium are both isotopes of hydrogen with mass numbers 2 and 3, the nuclear reaction expressed by Eq (11 2-1) is often written in the form†

$${}_1H^2 + {}_1H^2 \rightarrow {}_1H^3 + {}_1H^1 \qquad (11\ 2\text{-}2)$$

where ${}_1H^1$, ${}_1H^2$, and ${}_1H^3$ represent the nuclei of hydrogen (which is nothing but a proton), deuterium, and tritium, respectively The following deuterium–tritium fusion reaction‡

$$D + T \rightarrow \alpha(3\ 5\ \text{MeV}) + \text{neutron}\ (14\ 1\ \text{MeV}) \qquad (11\ 2\text{-}3)$$

is also of considerable importance as a possible source of thermonuclear power As indicated in Eq (11 2-3), the total energy liberated is about 17 6 MeV Tritium does not occur naturally, and one of the methods for producing it is to let the neutron [emitted in the D–T reaction—see Eq (11 2-3)] interact with lithium [see Eq (11 4-1)] Even though deuterium is available in abundant quantities (it constitutes about 0 015% of natural water), one expects to use the D–T reaction in a fusion reactor, because at $T \simeq 100$ million °K§, the D–T reaction is about 100 times more probable than the D–D reaction‖ and the energy released in the D–T reaction is about four times that in a D–D reaction [see Eqs (11 2-1) and (11 2-3)]

11 3. The Laser Energy Requirements

One of the difficulties associated with the fusion reaction is the requirement of a very high temperature for the fusion reactions to occur This is due to the fact that unless the nuclei have very high kinetic energies, the Coulomb repulsion will not allow them to come sufficiently close for fusion reactions to occur¶ The temperatures required are

† The nuclei are identified by symbols like ${}_{11}Na^{23}$ the subscript (which is usually omitted) represents the number of protons in the nucleus and the superscript represents the total number of nucleons in the nucleus Thus, ${}_{11}Na^{23}$ represents the sodium nucleus having 11 protons and 12 neutrons Similarly ${}_1H^3$ represents the tritium nucleus having 1 proton and 2 neutrons

‡ Equation (11 2-3) can also be written in the form

$${}_1H^2 + {}_1H^3 \rightarrow {}_2He^4 + {}_0n^1$$

§ Temperatures of the order of 100 million °K are required in fusion reactors see Section 11 3

‖ See e g , Booth *et al* (1976)

¶ This is in contrast to fission reactions which are induced by neutrons which carry no charge As such even at room temperatures there is a considerable probability for fission reactions to occur and hence it is relatively easy to construct a fission reactor It may be mentioned that in a hydrogen bomb (where the fusion reactions are responsible for the liberation of energy), a fission bomb is first exploded to create the high temperatures required for fusion reactions to occur

usually ~ 100 million °K, and at such high temperatures the matter is in a fully ionized state and its confinement poses a serious problem, matter in a fully ionized state is known as a plasma Thus, two major problems in thermonuclear fusion are (i) heating of plasmas to very high temperatures and (ii) confinement of plasmas for times long enough for substantial fusion reactions to occur† For example, for the deuterium–tritium reaction [see Eq (11 2-2)] at 10 keV (≈ 100 million °K), for the fusion output energy to exceed the input energy required to heat the plasma, one must have

$$n\tau \gtrsim 10^{14}\ \text{cm}^{-3}\ \text{sec} \tag{11 3-1}$$

where n represents the plasma density and τ is the confinement time Equation (11 3-1) is known as the Lawson's criterion (see, e g, Ribe, 1975) For $n \sim 10^{15}$ ions/cm³, τ must be $\geq 0\ 1$ sec Although the plasma has not yet been confined for such long times, the Russian device (known as Tokamak), using magnetic confinement, has come close to the conditions where the above equality is satisfied

With the availability of intense laser pulses, a new idea of fusion systems has emerged The idea is essentially compressing, heating, and confining the thermonuclear material by inertial forces which are generated when an intense laser pulse interacts with the thermonuclear material, which is usually in the form of a solid pellet In such a confinement it is not necessary to have a magnetic field

For laser-induced fusion systems, instead of using the parameter $n\tau$, it is more useful to use the parameter ρR, where ρ and R represent the density of the fuel and the fuel radius, respectively It has been shown (see, for example, Booth *et al*, 1976, Ribe, 1975) that if f represents the fractional burn-up of the fuel then

$$f \simeq \frac{\rho R}{6 + \rho R} \tag{11 3-2}$$

where ρ is measured in g/cm³ and R is measured in cm For $f \simeq 0\ 05$ (i e, 5% burn-up of the fuel) ρR must be about 0 3 g/cm², the higher the value of ρR, the greater will be the fractional burn-up of the fuel Further, if the mass of D–T pellet is M gm then the total fusion energy

† In the sun (the energy of which is due to thermonuclear reactions), the plasma has a temperature of $\gtrsim 10$ million °K and it is believed that the confinement is due to the gravitational forces

released (in joules) would be given by†

$$E_{\text{output}} = 4.2 \times 10^{11} fM \text{ (J)} \tag{11.3-3}$$

Obviously

$$M = \frac{4\pi}{3} R^3 \rho = \frac{4\pi}{3} \frac{1}{\rho^2} (\rho R)^3 \tag{11.3-4}$$

Also, to heat the D–T pellet to temperatures ($\simeq$ 100 million °K) at which fusion reactions will occur with high probability, the laser energy required would be‡

$$E_{\text{laser}} \simeq 4 \times 10^8 \frac{M}{\epsilon}$$

or

$$E_{\text{laser}} \simeq 4 \times 10^8 \frac{4\pi}{3} \frac{1}{\rho^2 \epsilon} (\rho R)^3 \tag{11.3-5}$$

where ϵ represents the fraction of laser energy used for heating the pellet; usually $\epsilon \sim 0.1$ Thus the yield ratio, Y, is given by

$$Y = \frac{E_{\text{output}}}{E_{\text{laser}}} \simeq \frac{4 \times 10^{11} fM\epsilon}{4 \times 10^8 M} = 10^3 \epsilon f \tag{11.3-6}$$

† Since the masses of D and T nuclei are in the ratio of 2 3, the number of D nuclei will be

$$\frac{2M}{5} \frac{1}{M_d} = \frac{2M}{5} \frac{1}{2 \times 1.66 \times 10^{-24}}$$

where M_d ($\simeq 2 \times 1.66 \times 10^{-23}$ g) represents the mass of the deuteron we have assumed equal numbers of D and T nuclei in the pellet The energy released in a D–T reaction is 17.6 MeV [see Eq (11.2-3)] and an additional 4.8 MeV is released when the neutron is absorbed by the lithium atoms in the blanket [see Eq (11.4-1)] resulting in a net energy release of about 22 MeV Thus the output energy would be

$$E_{\text{output}} \simeq f \times \frac{2M}{5} \times \frac{22 \times 1.6 \times 10^{-6}}{2 \times 1.66 \times 10^{-24}}$$

$$\simeq 4.2 \times 10^{11} fM \text{ (J)}$$

‡ For the kinetic energies to be about 10 KeV ($\simeq$ 100 million °K), the energy imparted would be

$$2 \times \left(\frac{2M \times 10 \times 10^3 \times 1.6 \times 10^{-19}}{5 \times 2 \times 1.66 \times 10^{-24}} \right) \simeq 4 \times 10^8 M \text{ (J)}$$

where M is in grams the quantity inside the parentheses is the energy imparted to the deuteron and the factor of 2 outside the parentheses is due to the fact that an equal energy has also to be imparted to tritium nuclei

or,

$$Y \simeq 100 \frac{\rho R}{6 + \rho R} \tag{11 3-7}$$

where we have assumed $\epsilon \simeq 0\ 1$ Clearly, for $Y > 1$, $\rho R \geqslant 0\ 1\ \mathrm{g/cm^2}$ Thus for a sizable burn-up and for a reasonable yield one should at least have $\rho R \simeq 0\ 2\ \mathrm{g/cm^2}$ Now, for normal (D–T) solid, $\rho \simeq 0\ 2\ \mathrm{g/cm^3}$, using Eq (11 3-5) one obtains

$$E_{\text{laser}} = 4 \times 10^8 \times \frac{4\pi}{3} \frac{(0\ 2)^3}{(0\ 2)^2 \times 0\ 1} \simeq 3 \times 10^9\ \mathrm{J} \tag{11 3-8}$$

which is indeed very high However, as is obvious from Eq (11 3-5), for a given value of ρR, the laser energy requirement can be significantly decreased (to the megajoule range) by increasing the value of ρ, this can be achieved by compressions to 10^3 to 10^4 times the normal solid density At such high densities, the α particles that are produced in the reaction, give up most of their energy to the unburned fuel before leaving the pellet, this leads to an increased fractional burning of fuel For a typical calculation reported by Ribe (1975), for $\rho R \simeq 3\ \mathrm{g/cm^2}$ [corresponding to 30% burn-up—see Eq (11 3-2)] and an ignition temperature of 10^4 eV ($\simeq 100$ million °K), one obtains using Eq (11 3-5)

$$E_{\text{laser}} = 4 \times 10^8 \times \frac{4\pi}{3} \frac{(3)^3}{(0\ 2)^2 \times 0\ 4} \simeq 2\ 5 \times 10^{12}\ \mathrm{J} \tag{11 3-9}$$

where we have assumed a normal D–T solid of density $0\ 2\ \mathrm{g/cm^3}$ and $\epsilon = 0\ 4$ On the other hand for $\rho = 1000\ \mathrm{g/cm^3}$, one would obtain (for the same value of ρR),

$$E_{\text{laser}} \simeq 1\ 2 \times 10^5\ \mathrm{J} \tag{11 3-10}$$

Such high-energy laser pulses are expected to be achieved in the near future The above numbers correspond to a pellet mass of 110 μg whose radius (before compression) is about 0 05 cm, the compressed radius of this pellet is $\sim 0\ 003$ cm The fusion energy yield would be given by [see Eq (11 3-3)]

$$\begin{aligned} E_{\text{output}} &\simeq 4\ 2 \times 10^{11} \times 0\ 3 \times 100 \times 10^{-6} \\ &\simeq 14 \times 10^6\ \mathrm{J} = 14\ \mathrm{MJ} \end{aligned} \tag{11 3-11}$$

This corresponds to a yield of about 100

11.4. The Laser-Induced Fusion Reactor

In a laser-induced fusion reactor, we take for example a deuterium–tritium pellet in the form of a cryogenic solid where particle densities are $\sim 4 \times 10^{22}\,\text{cm}^{-3}$ and shine laser light from all directions (see Fig 11 1) Within a very short time, the outer surface of the pellet is heated considerably and gets converted into a very hot plasma ($T \sim 100$ million °K) This hot ablation layer expands into vacuum and as a reaction gives a push to the rest of the pellet in the opposite direction Thus, if a spherical pellet is irradiated from all sides, then a spherical implosion front travels towards the core For a deuterium–tritium plasma, with an incident intensity of $10^{17}\,\text{W/cm}^2$, one can get an inward pressure $\sim 10^{12}$ atmospheres† As the implosion front accelerates towards the center, it sets up a sequence of shock waves traveling inwards Such shock waves lead to a very high compression of the core and the fusion energy is released from high compression densities‡ along with the high temperature In order to obtain high compression densities, the time variation of the laser pulse has to be such that successive shock waves do not meet until they reach the center of the pellet It has been shown that the time

† The radiation pressure corresponding to an intensity of $10^{17}\,\text{W/cm}^2$ is only about 10^8 atm

‡ A compression ratio of a few thousand puts the laser energy requirement in the 10^5 J range (see Section 11 3)

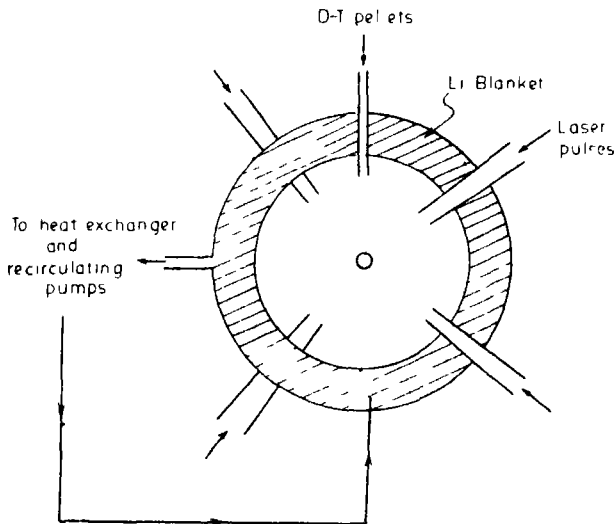



Fig 11 1 Schematic of a D–T fusion reactor with lithium blanket

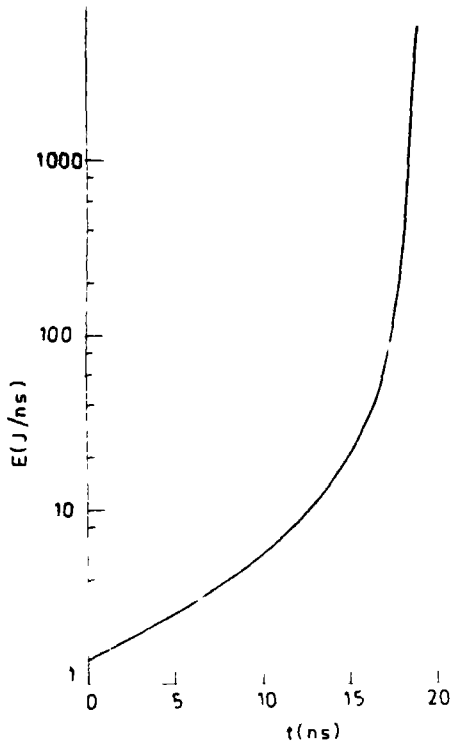



Fig. 11.2 Typical energy profile of the incident laser pulse for maximum compression in D–T pellets.

variation of the laser power should roughly be of the form $(t_0 - t)^{-2}$, where $t = t_0$ is the time when all the shocks reach the center. Thus if the pulse duration is about 1.0 nsec, about four-fifths of the energy goes in the last nanosecond and one-fifth in the first nine nanoseconds. A typical ideal energy profile for maximum compression in D–T pellets is shown in Fig. 11.2.

According to a computer experiment by Nuckolls and his co-workers (1972) a deuterium–tritium spherical pellet of radius 0.04 cm was irradiated by a 6×10^4 J pulse of 25 nsec duration† ($\lambda \sim 1\ \mu m$) compression densities as high as 1000 g/cm^3 were obtained and about 1.8×10^6 J of fusion energy was released in about 10^{-11} sec after the compression. Thus a multiplication by a factor of about 30 was observed.

We would like to mention that laser wavelengths $\sim 10\ \mu m$ may be too high for pellet heating. Detailed calculations show‡ that 1.06 μm

† The time variation of the incident laser pulse was assumed to be roughly of the form $(t_0 - t)^{-2}$, thus the power varied from about 10^{11} W at 10 nsec to 10[illegible] W at 15 nsec.
‡ See e.g. R. E. Kidder (1973).

radiation can heat a deuterium–tritium pellet to five times the temperature in one tenth the time as compared to 10 6 μm radiation heating the same pellet This and other facts suggest the use of 1-μm radiation for laser-induced fusion The 10 6-μm radiation corresponds to a CO_2 laser (see Section 9 6) and the 1 06-μm radiation corresponds to the

Table 11 1 [a] *Laser Fusion Facilities (Existing and under Production)*

Location[b]	Type	Number of beams	Total beam area (cm^2)	Maximum energy per beam	Peak power (TW)
USA					
LASL[c]	CO_2	8	9600	10 kJ	20
LASL[c]	CO_2	72	—	100 kJ	100–200
LASL[c]	CO_2	1	1200	400 J	0 4
KMS[c]	Nd glass	2	200	200 J	0 5
NRL[c]	Nd glass	2	70	300 J	0 2
LLL					
ARGUS[c]	Nd glass	2	600	1 kJ	4 0
SHIVA[c]	Nd glass	20	6300	10 kJ	30 0
NOVA	Nd glass	100	—	300–500 kJ	300
LLE					
GDL[c]	Nd glass	1	60	210 J	0 7
ZETA[c]	Nd glass	6	360	1 3 kJ	3–5
OMEGA	Nd glass	24	5500	10–14 kJ	30–40
USSR					
Lebedev					
UM1 35[c]	Nd glass	32	2560	10 kJ	5–10
DELPHIN	Nd glass	216	3430	13 kJ	10–15
UK					
Rutherford[c]	Nd glass	2	200	200 J	0 5
France					
LiMeil (Octal)	Nd glass	8	500	700 J	1 0
Japan					
Osaka	Nd glass	12			40

[a] Table adapted from the following sources Following foreign fusion Three countries *Opt Spectra* **13**(5) 30 (1979) Laser fusion and the energy challange A rundown on 1979 *Opt Spectra* **13**(11) 29 (1979) Shiva moves closer to laser fusion *Phys Today* **32**(11) 20 (1979) Rochester operates user facility for laser fusion *Phys Today* **32**(2) 17 (1979) Nova fusion laser *Laser Focus* **15**(7) 38 (1979)

[b] LASL—Los Alamos Scientific Laboratory Los Alamos New Mexico KMS—KMS Fusion Ann Arbor Michigan NRL—Naval Research Laboratory Washington DC LLL—Lawrence Livermore Laboratory Livermore California LLE—Laboratory for Laser Energetics Rochester New York

[c] Existing facilities

Fig. 11.3. A beam of infrared light from a Nd glass laser after passing through a 10-cm aperture potassium dihydrogen phosphate (KDP) crystal (at left) is halved in wavelength (and hence doubled in frequency), emerges as green light at 5320 Å, and is reflected by the mirror at the foreground into the target chamber at far right. The 12.8-mm-thick KDP crystals have yielded doubling efficiencies greater than 60%. The output power was 0.5 TW ($= 0.5 \times 10^{12}$ W), one of the highest powers in the visible region of the spectrum. (Photograph courtesy Thomas A. Leonard, KMS Fusion Inc.)

neodymium-doped glass lasers (see Section 9.5). Further, neodymium-doped glass lasers are capable of delivering a high laser energy within a short time. Energies ~200 J in times ~0.1 nsec have already been achieved (see Table 11.1). One of the major drawbacks of the neodymium-doped glass laser system is the fact that the efficiency (defined as the ratio of the laser energy output to the electrical energy input) is extremely low (~0.2%). On the other hand, CO_2 lasers† (which operate at 10.6 μm) have efficiencies in the range 5% to 7%. Although it is difficult to predict the specific laser systems which will be in operation in a laser fusion reactor, one does expect that very soon the laser technology would be sufficiently developed to meet the requirements.

† For details of other kinds of lasers used in fusion see, e.g., Booth *et al.* (1976).

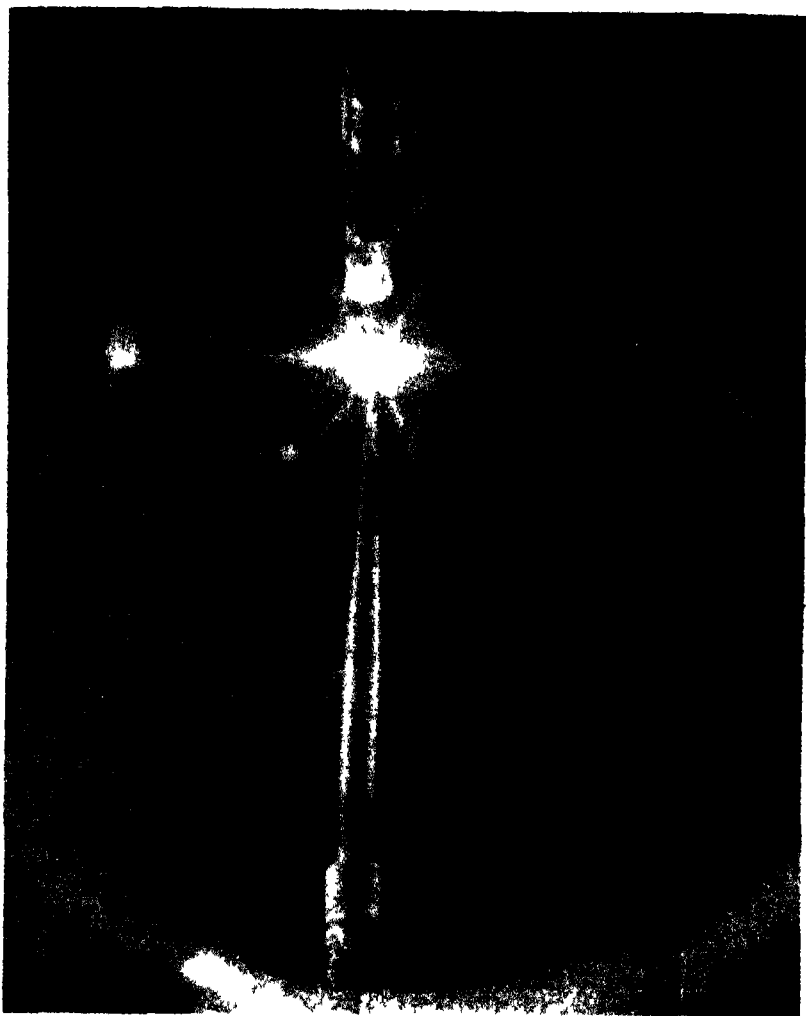



Fig 11 4 Laser irradiation of a cryogenic target which is a hollow spherical glass shell 51 μm in diameter, with a 0 7 μm wall, containing 1 3 ng of deuterium–tritium condensed in a liquid layer on the inside surface of the shell Two X-ray pinhole cameras are seen projecting towards the target from the top of the chamber The targets produced 7×10^7 neutrons (Photograph courtesy Thomas A Leonard, KMS Fusion Inc)

Figures 11 3 and 11 4 show some photographs of the laser fusion experiments carried out at KMS Fusion Inc , USA Figure 11 3 shows a beam of light emerging from a neodymium glass laser after passage through a 10-cm aperture potassium dihydrogen phosphate (KDP) crystal (at left) which doubles the frequency (harmonic generation) and hence halves the wavelength from 1 064 μm (infrared) to 5320 Å (green), which then enters the target chamber shown at the extreme right Output powers of 0 5 TW (1 TW = 10^{12} W) were measured at the output of the KDP crystals Figure 11 4 shows a photograph taken during the laser irradiation of the cryogenic target which was a hollow spherical glass shell 51 μm in diameter with a 0 7 μm wall containing 1 3 nanogram of deuterium–tritium condensed into a liquid layer on the inside surface of the shell The targets produced 7×10^7 neutrons on irradiation

We end this chapter by briefly describing the laser fusion reactor (usually abbreviated as LFR) A simplified block diagram of the electric generating station is shown in Fig 11 5 The reactor would roughly consist of a high-vacuum enclosure at the center of which deuterium–tritium pellets are dropped at regular intervals of time As soon as the pellet reaches the center of the chamber it is irradiated by synchronized pulses from an array of focused laser beams from all directions (see Fig 11 4) The fusion energy released is absorbed by the walls of the chamber consequently the walls get heated up, which may be used for running a steam turbine For a commercial power station, if 100 pellets are allowed to explode per second and if in each explosion, an energy of about 10^8 J is released, one would have a 10 GW ($= 10^{10}$ W) power station

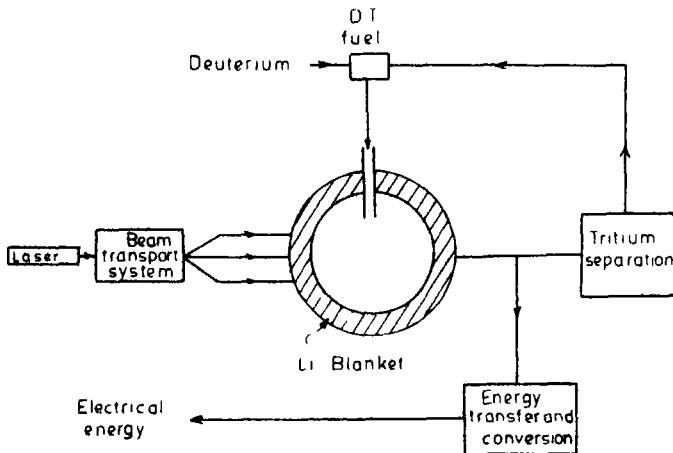
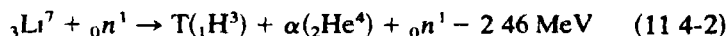



Fig 11 5 A simplified block diagram of a laser-fusion electric generation system

One of the most important aspects of any fusion reactor is the production of tritium Since it is not available in nature, it has to be produced through nuclear reactions This is achieved by placing lithium† (or its compounds) in a blanket surrounding the reactor chamber The neutron emitted in the fusion reaction [see Eq (11 2-3)] is absorbed by a lithium nucleus to give rise to tritium according to either of the following reactions

$$_3Li^6 + {_0n^1} \rightarrow T(_1H^3) + \alpha(_2He^4) + 4\ 79\,\mathrm{MeV} \qquad (11\ 4\text{-}1)$$

$$_3Li^7 + {_0n^1} \rightarrow T(_1H^3) + \alpha(_2He^4) + {_0n^1} - 2\ 46\,\mathrm{MeV} \qquad (11\ 4\text{-}2)$$

Notice that in the second reaction, the neutron appears on the right-hand side also, which can again interact with a lithium nucleus to produce tritium‡ The conceptual designs consist of liquid lithium being contained between two structural shells which enclose the reactor cavity Liquid lithium will also be responsible for the removal of heat from the reactor and the running of a steam turbine The heat exchangers and the lithium processing equipment for separation of tritium are expected to be adjacent to the reactor (see Fig 11 5)

We conclude by noting that there are many practical difficulties associated with the laser-driven fusion system, such as (a) delivery of a substantial part of the laser energy to the fuel before heating the fuel (which would, in effect, reduce the compression achieved) and before the shock wave disperses the fuel mass, (b) building of high-power lasers to a stage when the output energy from the system exceeds the input energy, (c) design of complex targets and reliable production of such targets with extremely good surface finish, as surface irregularities of more than 1% of the thickness of the wall seem to yield very unstable compressions of thick shells, etc Also, projects are underway to study the feasibility of using particle beams like electron beams and ion beams as fusion drivers It is much beyond the scope of the present book to go into the details of the various difficulties, the interested reader may look up Brueckner and Jorna (1974), Post (1973), and Stickley (1978) for further details

† Lithium is quite abundant and has good heat transfer properties

‡ In addition neutron multiplication will occur in the blanket through (n $2n$) reactions these neutrons would further produce more tritium

12

Lightwave Communications

12 1 Introduction

One of the most important and exciting applications of lasers lies in the field of communications Since optical frequencies are extremely large ($\sim 10^{15}$ Hz), as compared to the conventional radio waves ($\sim 10^{6}$ Hz) and microwaves ($\sim 10^{9}$ Hz), a light beam acting as a carrier wave is capable of carrying far more information in comparison to radio waves and microwaves (see Section 12 2) It is expected that in the not too distant future, the demand for flow of information traffic will be so high that only a lightwave would be able to cope with it

Soon after the discovery of the laser, some preliminary experiments on propagation of information-carrying lightwaves through the open atmosphere were carried out, but it was realized that because of the vagaries of the terrestrial atmosphere—e g , rain, fog, etc —in order to have an efficient and dependable communication system, one would require a guiding medium in which the information-carrying lightwaves could be transmitted Even under clear atmospheric conditions, because of the turbulent nature of the atmosphere, the lightwave after passing some distance through the atmosphere, suffers from a low-frequency noise modulation In addition, because of atmospheric turbulence, it may even so happen that the beam completely misses the receiver These effects are obviously absent in outer space, where there is no atmosphere Thus two spaceships can communicate using lightwaves although here there is an additional problem of pointing and locating the highly directional and thin optical beam by the receiver Thus for terrestrial applications the necessity of a proper guiding medium was recognized and extensive work began in this direction

Initially some attempts were made to send light beams through conduits, which could be evacuated and fitted with lenses and mirrors at periodic distances for beam guidance, or could be filled with a gas in which a temperature gradient could be set up In the latter approach, the

temperature gradient would lead to a transverse refractive index gradient which would, in turn, provide a guidance mechanism for the lightwave† However, it was soon realized that such schemes would be extremely cumbersome and unwieldy in practice and would also be quite unreliable

Another scheme, which was already known at that time, was the transmission of lightwaves through glass fibers employing the phenomenon of total internal reflection Such a scheme was not regarded as practical for long-distance communication because even the best-quality optical glass was known to have extremely high transmission loss (≥ 1000 dB/km)‡ With such huge losses, the use of glass fibers as a communication medium would be highly impractical The losses in the fibers arise primarily from absorption, scattering, and imperfections in the fiber For example, glass usually contains metallic ions like Cr^{3+}, Cu^{2+}, Fe^{2+}, etc , and the hydroxyl ion OH^- (due to the presence of water) in various concentrations and even an impurity of 1 part in 10^9 of Fe^{2+} or Cr^{3+} ions can cause a loss of ~ 1 dB/km Thus, it becomes clear that extremely high purity is required in the manufacture of these fibers It was in the year 1966 that C Kao and G A Hockham in England, proposed, for the first time, the possibility of using glass fibers for communications by reducing the loss considerably In 1970, Corning Laboratories in USA announced the first practical realization of optical fiber waveguides with losses as low as 20 dB/km at the He–Ne laser wavelength Since then the loss has been declining steadily and presently fibers with losses of a few dB/km are being routinely fabricated In fact, fibers having a loss as low as 0 2 dB/km at a wavelength of 1 55 μm have recently been fabricated (Miya *et al*, 1979) such low-loss fibers are highly attractive for extremely long communication links

It may be mentioned here that as a light source one could use either an incoherent source like a light-emitting diode (LED) or a coherent source like a semiconductor laser, the former being much less expensive However, since the spectral width of a laser is much smaller than an LED, the information capacity of a system employing lasers is much

† These are known as gas lenses for further details one may look up Sodha and Ghatak (1977)

‡ If an input power P_1 results in an output P_2 then the loss in decibels (dB) is given by

$$10 \log_{10} \frac{P_1}{P_2}$$

Thus if the output power is only half of the input power then the loss is $10 \log 2 \approx 3$ dB Similarly a loss of 30 dB corresponds to an output power to input power ratio of 1/1000 Thus a loss of 1000 dB/km amounts to the fact that when the light beam travels through a kilometer length of the fiber only a negligible amount reaches the output end

higher than one using LEDs (see Section 12 3) In addition, laser light can also be more efficiently coupled to a fiber In fact, in the future super-high-bandwidth communication systems employing single-mode fibers, only a laser can be used since very little power can be launched into a single-mode fiber from an LED For more details on laser sources used for communication purposes, see, e g , Kressel *et al* (1976), Kressel (1979)

In addition to the widespread application of lasers in terrestrial communication (e g , interoffice, intercity, intracity, etc) laser beams can also be used for communicating between two satellites or between a satellite and ground Such a technique offers some important advantages for the military over the microwave system—namely, those of higher information capacity and also of immunity from jamming and interception At present the United States Air Force is going ahead with such a space laser communication project at an estimated cost of $30 million and the system is shortly expected to go in for demonstration The system would transmit up to 14 high-resolution television pictures simultaneously†

Lasers also offer unique advantages in communicating with a spacecraft at the time of its reentry to the atmosphere It is well known that when a spacecraft enters the atmosphere, there is a tremendous heating of the body of the spacecraft and a very high-temperature plasma is formed around the vehicle Because of the presence of the plasma, it becomes very difficult to communicate with the spacecraft using radio waves This is because the plasma acts as a conductor for the radio waves (whose frequencies are $\sim 10^6$ Hz) and consequently, these waves are not transmitted through the plasma layer This results in what is known as the 'blackout'' during the reentry This plasma is, however, transparent to optical frequencies ($\sim 10^{15}$ Hz) and a laser beam can be used to communicate from the spacecraft to the ground station

12.2 Large Information-Carrying Capacity of Lightwaves

In order to understand the potential of laser beams to carry a vast amount of information, we begin by considering the example of transmission of audio signals over a large distance When we speak, we produce sound waves which mostly vary in frequency from about 20 Hz to about 4000 Hz In order to transmit the sound signals over a large distance, the

† *Laser Focus*, June 1977, p 12

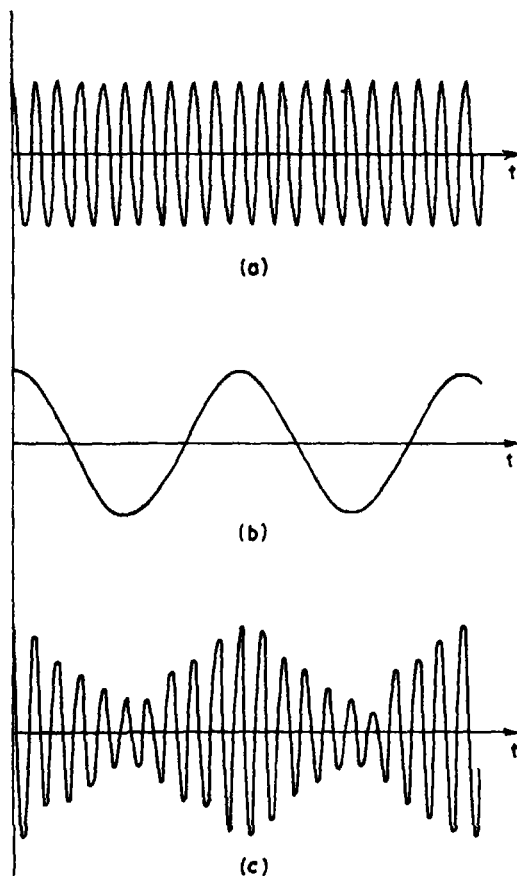



Fig 12 1 When the carrier wave shown in (a) is amplitude modulated by the modulating wave shown in (b), one obtains an amplitude-modulated wave as shown in (c)

audio signal is made to modulate a radio wave†, i e , some parameter of the radio wave like amplitude, frequency, or phase is made to change in a manner proportional to the audio signal Thus the modulated radio wave has coded into it the audio waves (which is the information to be sent) The radio wave here carries the information and hence is called a carrier wave The audio wave modulates the radio wave and hence is called the modulating wave

† Radio waves are electromagnetic waves having a much larger wavelength than that of light waves, and hence a much lower frequency Typical radio wave frequencies are in the range of 10^4 to 10^6 Hz

There are various ways of modulating the radio waves For example, one may modulate the amplitude of the radio wave in accordance with the signal Such a scheme results in what is known as amplitude modulation and is depicted in Fig 12 1 Similarly, instead of modulating the amplitude of the radio wave, one may modulate the frequency of the radio wave in accordance with the audio wave This results in frequency modulation and is shown in Fig 12 2

Let us first consider the amplitude-modulating scheme It can be shown that when a carrier wave of frequency ν_c is modulated with a modulating wave of frequency ν_m then the modulated wave consists of waves of frequencies $\nu_c - \nu_m$ and $\nu_c + \nu_m$ in addition to waves of frequency ν_c The waves at frequencies $\nu_c - \nu_m$ and $\nu_c + \nu_m$ are said to lie in the lower side band and upper side band, respectively In general, the modulating wave may consist of waves of frequencies from 0 to ν_m Thus on modulating, one obtains two bands of frequencies lying from $\nu_c - \nu_m$ to ν_c (the lower side band) and another lying between ν_c and $\nu_c + \nu_m$ (the upper side band) Since both the side bands contain the information, one may transmit only one of the side bands Such a system is called single-side-band transmission One can extract the information out of the signal by again "mixing" the received wave with the frequency

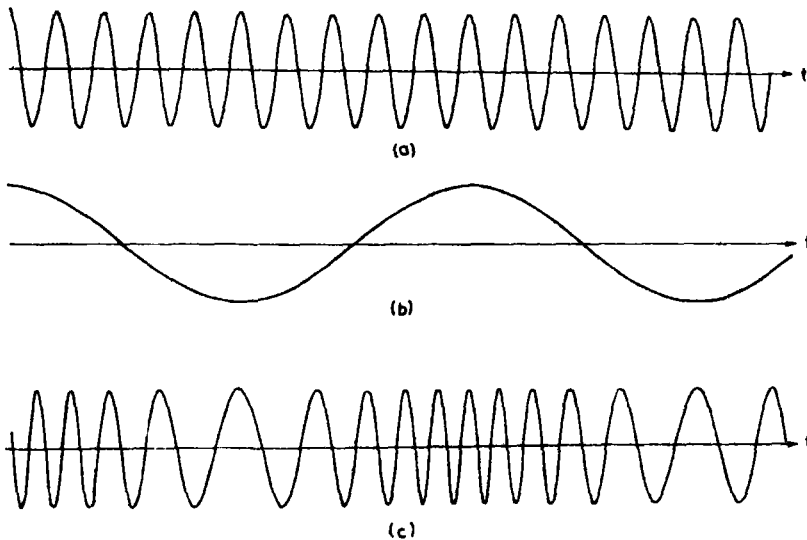



Fig 12 2 When the carrier wave shown in (a) is frequency modulated by the modulating wave shown in (b), one obtains a frequency-modulated wave as shown in (c)

ν_c Thus when the wave with frequency $\nu_c - \nu_m$ is mixed with a wave of frequency ν_c, then one would obtain waves of frequencies ν_m, ν_c and $2\nu_c - \nu_m$ One of these corresponds to ν_m, the modulating signal

Since the frequency of the speech signal may lie anywhere between 0 and 4000 Hz, it becomes clear that if the carrier wave has a frequency of, say, 100,000 Hz, then, for example, in the upper side band transmission, we must transmit frequencies between 100,000 and 104,000 Hz Thus, a band of at least 4000 Hz must be reserved for one speech signal Hence between carrier frequencies of 100,000 and 500,000 Hz, we can at most send (500,000 − 100,000)/4000 = 100 independent speech signals simultaneously Since the same bandwidth of 4000 Hz is required irrespective of the value of the carrier frequency, it becomes clear that in a higher carrier frequency channel between 10^{15} and 5×10^{15} Hz, one can send 10^{12} speech signals This is an enormous capacity indeed The bandwidth of 4000 Hz that we have considered is enough for intelligibility of speech But for music the bandwidth is about 20 kHz For television, the bandwidth is about 6 MHz Thus only a smaller number of television channels exist in the same carrier frequency band

The above method of sending simultaneously more than one signal along the same channel by assigning different frequency bands for each channel is referred to as frequency division multiplexing Notice that all the signals are overlapping in the time domain while they are nonoverlapping in the frequency domain The different signals are separated at the receiver by making use of filters which transmit only the frequency band corresponding to the signal of interest

The modulation system envisaged to be used in optical fiber communication is the pulse modulation system This is based on the sampling theorem, according to which a band-limited signal, i e , a signal which has no frequency components above a certain frequency (say ν_m), is *uniquely* determined by its values at uniform time intervals less than $1/2\nu_m$ (see Appendix A) Thus if the signal value is specified every $1/2\nu_m$ seconds, the sampling rate is given by $2\nu_m$ samples per second Consequently, instead of sending the complete signal continuously, one may only send the sampled values at a rate of $2\nu_m$ samples per second Hence, for example, since speech signals mostly contain frequencies up to 4 kHz, if the speech signal is specified at regular intervals of 1/8000 sec , i e , if the signal is sampled at the rate of 8000 samples per second, then the signal can be completely retrieved from these sampled values Thus, a simple system involving pulse modulation could be that one produces pulses of light at a rate of 8000 times per second with the amplitude of the pulse being proportional to the value of the signal at that time This is referred to as pulse amplitude modulation and is depicted in Fig 12 3

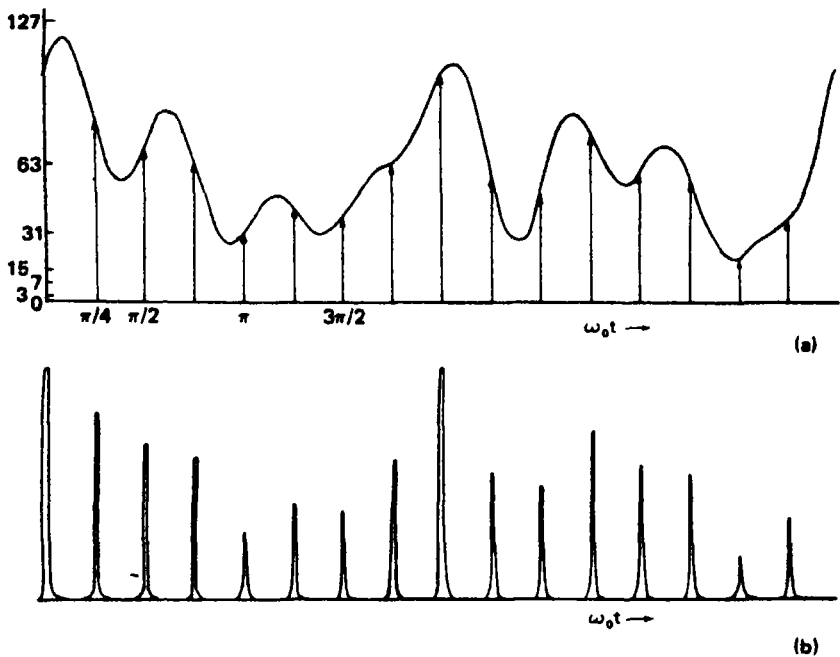



Fig 12 3 (a) The figure shows a time-varying signal given by Eq (12 2-1) whose maximum frequency component ν_m is $4\omega_0/2\pi$ Hence by the sampling theorem, the signal can be completely determined by sampling the signal at time intervals of $1/2\nu_m = \pi/4\omega_0$, which is shown as vertical lines (b) In the pulse amplitude modulation system, the amplitude of the signal at the sampled times is represented by the amplitude of the pulses

In addition to the pulse amplitude modulation technique, there exist other techniques of coding the signal value into the samples Thus one could modulate the width of the pulses in accordance with the signal value while keeping the pulse amplitude constant, this leads to pulse width modulation Similarly, one could alter the position of the pulse within the time interval leading to pulse position modulation

Another very important mode of transmission of the sampled values of the signal is called the pulse code modulation (PCM) In such a scheme, instead of varying the amplitude of a pulse in proportion to the value of the signal, the signal value is first converted into a code formed by a pattern of identical pulses In binary coding, the signal value is first converted to the binary code and then the ones and zeros in the code are represented by the presence or absence of a pulse Thus, in a coding system having seven levels, the signal value at a particular time is coded in the form of seven pulses, all the pulses being of the same height and same width

As an example, we consider a signal which is of the form

$$f(t) = 15[4 + \sin \omega_0 t + \sin(1\ 5\omega_0 t) + \cos(2\omega_0 t) + \cos(3\omega_0 t) + \sin(4\omega_0 t)] \quad (12\ 2\text{-}1)$$

where ω_0 is a constant The maximum frequency in the above signal is

$$\nu_m = \frac{4\omega_0}{2\pi} \quad (12\ 2\text{-}2)$$

Thus, according to sampling theorem, the sampling rate must be (at least)

$$2\nu_m = \frac{4\omega_0}{\pi} \text{ samples/sec} \quad (12\ 2\text{-}3)$$

Hence, the samples must be taken every $1/2\nu_m$ sec, i e , the time interval between two samples is

$$t = \frac{1}{2\nu_m} = \frac{\pi}{4\omega_0} \text{ sec} \quad (12\ 2\text{-}4)$$

In Figs 12 3a and 12 4a we have drawn the signal given by Eq (12 2-1) as a function of $\omega_0 t$, and as can be seen, the sample values are being taken every $\pi/4\omega_0$ sec Figure 12 3b corresponds to pulse amplitude modulation, where the amplitudes of the sampled pulses are equal to the amplitude of the signal

In order to consider the PCM system involving seven levels one must first code the signal value into a binary code of seven levels Figure 12 4a gives the signal value at the sampling times and the corresponding binary code In order to understand this, we note that the signal value of 105 corresponds to the binary equivalent 1101001 from the following logic

Division	Remainder
2 | 105	1
2 | 52	0
2 | 26	0
2 | 13	1
2 | 6	0
2 | 3	1
1	

The binary equivalent of 105 is formed by successively dividing by 2, then listing the final quotient and the remainders from bottom up Hence the binary equivalent of 105 is 1101001

The maximum signal value that can be coded into a binary equivalent (of seven levels) is $2^7 - 1 = 127$ Thus, the signal values can be coded into intervals of unity in a total amplitude of 127 This therefore corresponds to coding into equally spaced levels and becomes less accurate at low

Time, t	Signal Value	Binary Equivalent in Seven Levels	Pulse Code Waveform
0	105	1101001	
$\frac{\pi}{4\omega_0}$	84	1010100	
$\frac{\pi}{2\omega_0}$	70	1000110	
$\frac{3\pi}{4\omega_0}$	64	1000000	
$\frac{\pi}{\omega_0}$	30	0011110	
$\frac{5\pi}{\omega_0}$	44	0101100	

(a)

First Sample 105
1 1 0 1 0 0 1

Second Sample 84
1 0 1 0 1 0 0

Third Sample 70
1 0 0 0 1 1 0

Fourth Sample 64
1 0 0 0 0 0 0

Fifth Sample 30
0 0 1 1 1 1 0

Sixth Sample 44
0 1 0 1 1 0 0

(b)

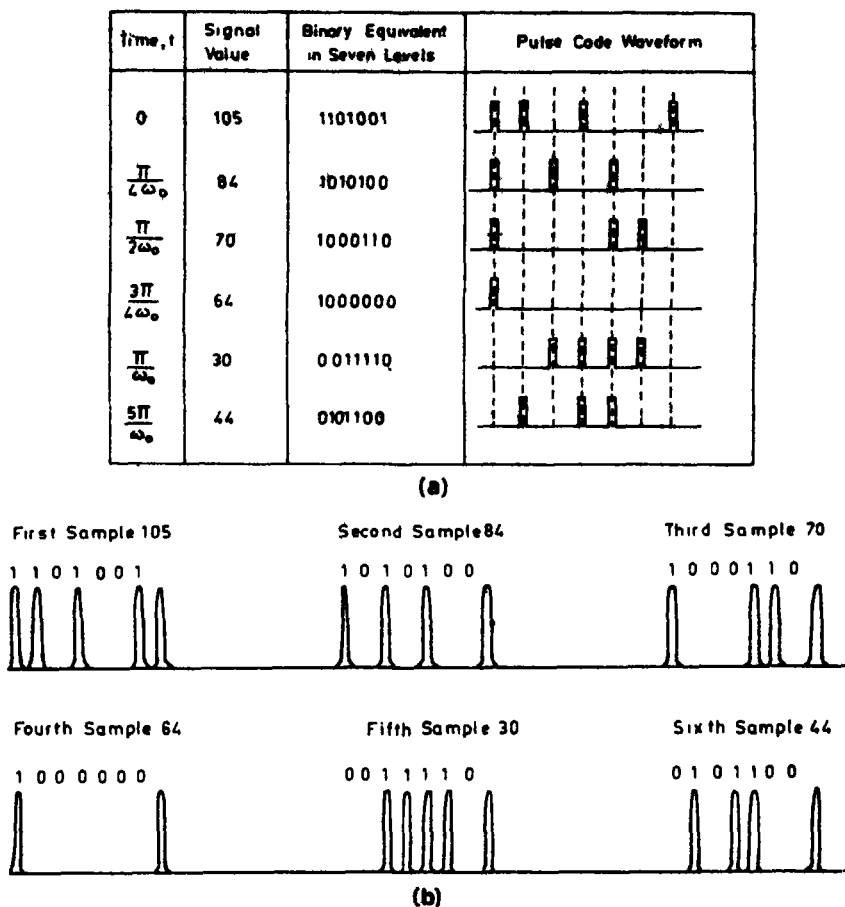



Fig 12 4 (a) Pulse code wave form of sample signal depicted in Fig 12 3a (b) In the pulse code modulation system (PCM) using binary coding involving seven levels, the sample values are first converted into the binary equivalent as shown in (a) Then the ones and zeros in the binary coding may be represented by the presence or absence of a pulse

amplitude values In order to overcome this, one often uses the logarithm scale coding system (see, e g , Cattermole, 1969)

In Fig 12 4b, the binary coded sampled values have been depicted in the form of pulses It may be noted here that in Fig 12 4b, at every sample, there are eight pulses instead of seven, the eighth pulse is required for synchronization of the receiver Thus for the above signal,

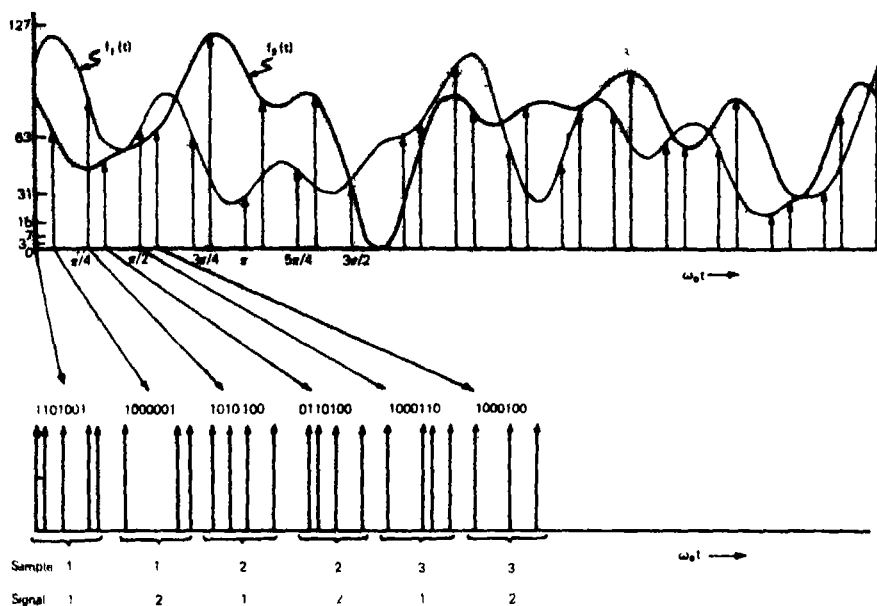



Fig 12 5 Curves 1 and 2 correspond to two different signals to be transmitted along the same channel simultaneously Each signal is first sampled, then coded into a binary code and then they are intermingled so that the pulses corresponding to the two signals appear at different times The interspaced series of pulses is then sent along the channel Such a scheme in which the different signals occupy different time intervals is called time division multiplexing

there would be $8 \times 4\omega_0/\pi = 32\omega_0/\pi$ pulses per second, the first factor of 8 corresponds to the fact that each sample is being replaced by eight pulses and the second factor corresponds to the sampling rate

Now coming back to the case of speech signal, since the maximum frequency is 4000 Hz the sampling rate should be 8000 samples per second It each sample is replaced by eight pulses, one would have $8 \times 8000 = 64\,000$ pulses per second, which is also written as 64,000 bits/sec[†] or 64 Kbit/sec

Hence, corresponding to each signal, one would have a series of pulses appearing 64,000 times per second Simultaneous transmission of many signals is done by intermingling the sample values of the various signals such that the pulses of different signals propagate at different times Figure 12 5 shows how two signals can be sent through the same channel on a time-sharing basis Such a scheme of multiplexing is known as time division multiplexing (TDM) In such a scheme all the signals

[†] The word *bit* is derived from binary digit

occupy the same frequency band but are separated in time This is in contrast to frequency division multiplexing in which the various signals occupy the same region in the time domain while they occupy different regions in frequency domain

The number of independent messages that can be sent simultaneously in a system employing TDM would be determined by the condition that the pulses at the output be well resolved in time so that information can be retrieved back In Section 12 3 1 we briefly discuss the phenomenon of pulse dispersion which leads to a broadening of the pulse as it propagates through the transmission medium This broadening ultimately would limit the information capacity of the system

If one compares the pulse amplitude modulation (PAM) and pulse code modulation (PCM) systems, it is apparent that the PCM system requires more time interval for the transmission of the same information since corresponding to each pulse in the PAM system, one has 8 pulses in the PCM system This disadvantage of the PCM system is offset by the fact that a system employing PCM is more immune to noise and interference effects This is because in the PCM system the receiver has only to detect the presence or absence of a pulse regardless of its amplitude value and width Thus all external factors which tend to distort the amplitude or the shape of the pulses have no effect on a PCM system, of course, the distortion should not be so much as to make the pulses unresolvable In addition to the above, because of the digital nature of PCM, information from a variety of sources—voice, video, etc —can be multiplexed for time sharing

It can be seen from the above that if the pulses can be made very narrow and still are resolvable at the output end of the transmitting channel, then the information capacity of the system would be much higher If the pulses are of infinitesimal width (i e , the sampling is by impulses) then the bandwidth of the sampled signal can be shown to be infinite On the other hand if the pulses are of finite width, the resulting sampled signal has negligible energy content at higher frequencies In fact, the bandwidth required for transmission becomes smaller with increase in pulse width It therefore appears that sampling with finite width pulses is superior to impulse sampling since it requires a smaller bandwidth But if the pulses have a finite width then the sampled signal requires a longer time interval for transmission and only a smaller number of signals can be transmitted simultaneously on a time-sharing basis

At optical frequencies, even a pulse of duration as small as 1 psec ($= 10^{-12}$ sec) is 1000 wavelengths long and has a bandwidth of $\sim 10^{12}$ Hz, which is much smaller than the carrier frequency of $\sim 10^{15}$ Hz As an example let us consider transmission through an optical fiber which

exhibits a pulse dispersion of ~0 1 nsec per kilometer of propagati (1 nsec = 10^{-9} sec) If we consider a communication link between tv points separated by 5 km and assume that the losses of the fiber are sm: so that no repeater is required between the two terminals, then in propagation through 5 km the pulse dispersion would be about 0 5 nse Hence, in order that the various pulses be resolvable at the output en the time interval between two pulses must be about 5 nsec Hence o can send about $1/5 \times 10^{-9}$ or 2×10^{8} pulses per second Thus tl maximum bit rate through such a system would be 2×10^{8} bits/sec (200 Mbit/sec (1 Mbit = 10^{6} bits) Since one speech signal requires tran mission at a rate of 64 Kbit/sec, one may send at most $2 \times 10^{8}/64$ $10^{3} \approx 3000$ independent speech signals simultaneously through the opt cal fiber

Field trials are underway in various countries around the wor including the USA, Japan, Europe, India, etc, to evaluate tl performance of fiber optic communication systems Typical bit rates in tl systems undergoing field trials are 6 3 Mbit/sec, 8 44 Mbit/sec, ar 44 7 Mbit/sec carrying, respectively, 96, 120, and 672 voice signals Tl results are very encouraging and it is expected that soon one would hav optical fiber communication links carrying commercial traffic (For mo details, see e g, Gallawa, 1979, Shimada, 1979, Midwinter, 1979b)

The field trials mentioned above are in the source wavelength aroun 0 8 μm at which GaAlAs laser diodes and LEDs operate Recent r searches in optical fibers have revealed a very interesting region (operation around the 1 3–1 6 μm wavelength in which it is found th both the loss and dispersion effects are much less than in the 0 8-μ wavelength region In fact a loss as low as 0 2 dB/km has been measure at a wavelength of 1 55 μm (Miya *et al*, 1979) Extensive researc programs are underway around the world for fabrication of sources ar detectors in the new wavelength range, sources fabricated using InGaAs structures seem to be the most promising at present Coupling of th operating wavelength region with low-dispersion single-mode fibers a expected to yield systems possessing superhigh bandwidth in excess (100 GHz (Gloge, 1979)

12.3. Components of a Lightwave Communication System

12 3 1. The Optical Fiber

As discussed earlier, the atmosphere cannot be used as a transmi sion channel for terrestrial communications using light beams The mo

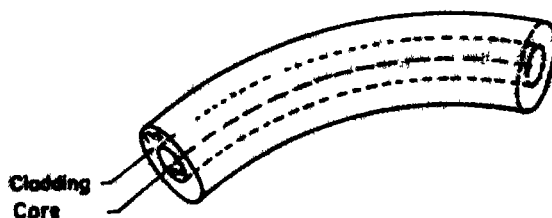



Fig. 12 6 A typical optical fiber consisting of a transparent material of refractive index n_1 and surrounded by a cladding of a slightly lower refractive index n_2 Typical dimensions of the core and cladding are 50 and 125 μm, respectively Most of the propagating energy is confined to the core region and the field decays exponentially in the cladding.

promising channel is the optical fiber waveguide An optical fiber† essentially consists of a central transparent region called the core which is surrounded by a region of lower refractive index called the cladding (see Figs 12 6 and 12 7) The core could either be homogeneous or could have a gradient in refractive index with the refractive index decreasing away from the center of the core In the former type of fibers, also referred to as homogeneous core fibers (see Fig 12 8), the guidance of light occurs through the phenomenon of total internal reflection at the core–cladding interface In the latter type of fibers, also referred to as graded-index fibers (see Fig 12 9) the guidance of light occurs through a continuous refraction of light rays towards the center of the core

In an optical waveguide, there exist specific field distributions which propagate without changing their form and with a definite phase and group velocity These field configurations are referred to as the modes of the optical waveguide These modes are characterized by different propagation constants and different group velocities In a multimode waveguide, there exists a large number of these propagating modes while in a single-mode waveguide there exists only one mode of propagation Each mode has most of the energy inside the core, but due to the evanescent fields outside the core, a part of the energy is also traveling in the cladding By making the cladding sufficiently thick, the fields of the mode at the cladding–air boundary can be made small, thus making it easy to handle and support these fibers without causing much disturbance to the modes

As already discussed, in a fiber optic communication system the information is coded in the form of discrete pulses which are transmitted

† The various techniques for the fabrication of low-loss optical fibers has been given in a recent review by Pal (1979) The propagation characteristics are discussed in many places, see, e.g., Barnoski (1976), Ghatak and Thyagarajan (1980), Midwinter (1979a)

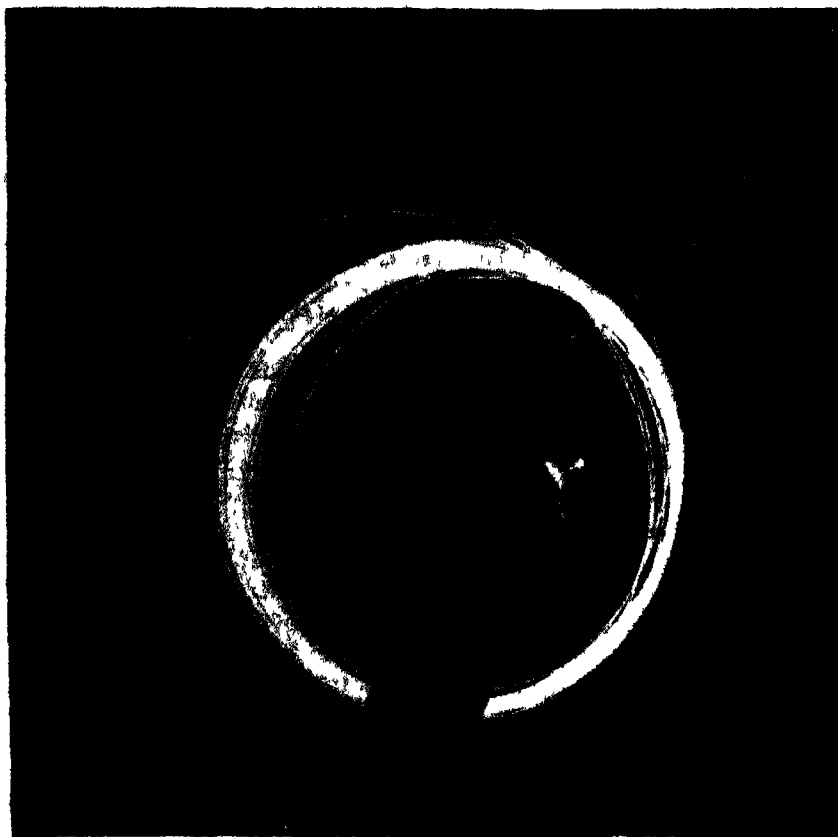



Fig 12 7 A long and thin optical fiber carrying a light beam (After Chynoweth 1976 photograph courtesy Dr A G Chynoweth)

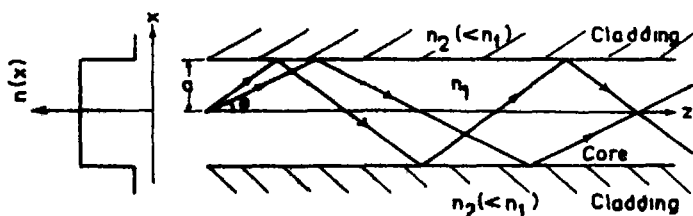



Fig 12 8 A homogeneous core optical fiber in which the refractive index in the core is constant. Light rays impinging on the core-cladding interface at an angle greater than the critical angle are trapped inside the core of the waveguide In such a fiber rays traveling at larger angles to the axis have to traverse a larger path and hence take a longer time than those rays which propagate with lesser angles to the axis This leads to a substantial amount of broadening in a pulse propagating through the fiber

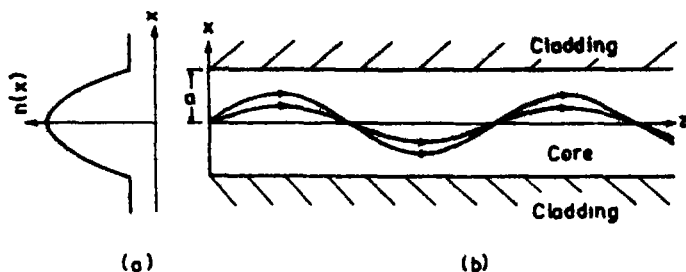



Fig 12 9 A graded index optical fiber in which the refractive index in the core decreases continuously away from the axis (a) Shows a typical variation of refractive index across the core of the optical fiber (b) Light rays in such a fiber are trapped by a continuous refraction towards the center of the core which leads to a periodic focusing of the rays In such a fiber, even though rays making larger angles with the axis traverse a longer path length, they do so in a region with a lower refractive index and hence a higher speed of propagation this leads to a smaller value of pulse dispersion in such fibers as compared to homogeneous core fibers

through the fiber The information capacity of the system will be determined by the number of pulses that can be sent per unit time For the information to be retrieved at the output end, the various pulses must be well resolved in time In an optical fiber due to various factors like the differences in group velocity between the different modes and the dependence of the propagation constant of a mode on wavelength, a pulse of light broadens as it propagates through the fiber Hence, even though two pulses may be well resolved at the input end, because of broadening of the pulses they may not be so at the output (see Fig 12 10) In such a case no information can be retrieved at the output Thus for a given broadening, the pulses have to be separated by a minimum time interval which would determine the ultimate information-carrying capacity of the system

When a pulse of radiation is injected into a fiber, it excites the various modes of the fiber Since each mode propagates with, in general, a different characteristic group velocity, the incident pulse of light broadens as it propagates through the fiber This is referred to as intermodal broadening When the fiber can carry only one propagating mode, i e , in a single-mode fiber, this broadening is absent, but due to the dependence of the propagation constant on wavelength, there is still some broadening, this is referred to as intramodal broadening Both intermodal and intramodal dispersions arise as a result of (a) waveguide effects and (b) material effects, the latter is due to the finite bandwidth of the source and the fact that at different wavelengths the refractive indices are different It may be mentioned here that since lasers have much smaller

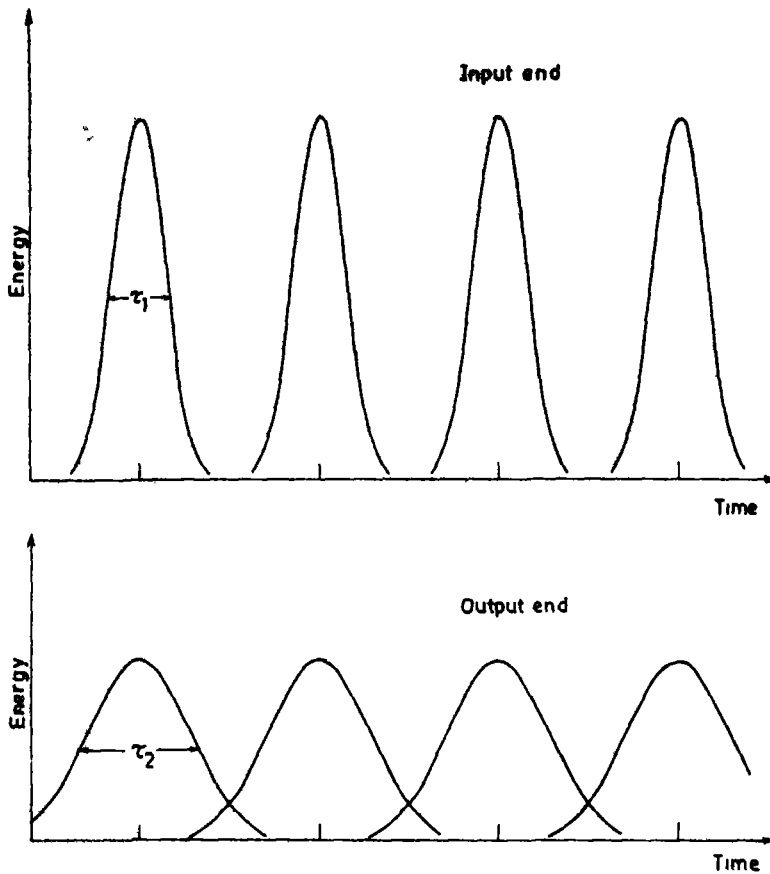



Fig 12 10 A series of pulses each of width τ_1 (at the input end of the fiber) after transmission through the fiber emerges as a series of pulses of width $\tau_2(>\tau_1)$ The broadening in the pulses is caused by the different group velocities of the various modes and by the dependence of the propagation constant on wavelength If the broadening is large, then adjacent pulses will overlap at the output end and may not be resolvable Thus, pulse broadening determines the minimum separation between adjacent pulses, which in turn determines the maximum information-carrying capacity of the optical fiber

spectral width as compared to light-emitting diodes, the material dispersion is much lower in a system employing lasers as compared to one using LEDs For example, with an LED, the pulse broadening due to material dispersion may be ~4 nsec/km, whereas with a laser this would be less than 0 2 nsec/km

The broadening of a pulse of light as it propagates through an optical fiber can also be visualized by using the concept of geometrical optics

When a pulse of light is injected into a homogeneous core optical fiber, it excites rays traveling at different angles with the axis As can be seen from Fig 12.8, since rays making larger angles with the axis have to traverse a longer optical path length, they take a longer time to reach the output end Consequently the pulse of light broadens as it propagates through the fiber In contrast, in a graded index fiber, even though rays making larger angles with the axis have to traverse longer path lengths they do so in a medium with a lower value of refractive index (see Fig 12 9) Thus the longer path length can be partially compensated by propagation at a higher velocity Hence the broadening of a pulse must be much lower in a graded index fiber as compared to a homogeneous core fiber In fact this is indeed the case, and for high-bandwidth applications, graded index fibers are more suitable than homogeneous core fibers

It may be mentioned here that in optical fibers having very small core radii and small index difference between the core and cladding, it can be so arranged that only one mode of propagation exists in the fiber (Ghatak and Thyagarajan, 1980) Such fibers are therefore referred to as single-mode fibers Because of the presence of just one mode, the dispersion in these fibers is very small and is only due to intramodal broadening Such fibers are indeed expected to be used in future super-high-bandwidth systems

In addition to the extremely large information-carrying capacity of a system using lightwaves, communication or transmission through optical fibers has several other additional advantages over the conventional metallic systems like the coaxial cable, etc

(i) Because of the extremely low transmission loss of practically available fibers, one can have much greater distance between repeater stations, resulting in substantial cost savings For example, Table 12 1

Table 12 1 [a] Typical Comparison of a Coaxial and a Fiber Optic System for Defense Applications

Item	CX-11230 system		Fiber optic	
Range (km)	8	64	8	64
Data rate (Mbit/sec)	19 6	2 3	19 6	2 3
Repeater	19	39	0	7
Cable cost ($)	7 000	56,000	9000	72,000
Repeater cost ($)	15 000	36,000	0	5,600
Total link cost ($)	22,000	92,000	9000	77,600
System weight (kg)	1,100	8,700	280	1,900
Transportation requirement		4 2½-ton trucks		1 2½-ton truck

[a] After Steensma and Mondrick (1979)

shows a comparison between a coaxial system and a fiber optic system typically for defense applications

(ii) Optical fibers are typically about a 100 μm in diameter and are basically made of silica or glass This results in a heavy reduction in weight and volume of space required, which is an important consideration for laying in already crowded available conduits This saving in weight and volume is also important for shipboard applications and data handling using optical fibers in aircrafts

(iii) Optical fibers are immune to electromagnetic interference and there is no cross talk This is an important consideration for secure communications in defense

(iv) Optical fibers can be used in explosive as well as high-voltage environments due to the absence of any hazard due to short circuits, etc

In addition to the main application in telecommunications, optical fibers are also expected to play an important role in computer links, space vehicles, industrial automation and process control, etc In fact, recently optical fibers have been used to carry data and control information within big fusion lasers at Lawrence Livermore Laboratory and Los Alamos Scientific Laboratory† and also for monitoring underground nuclear explosions at the Nevada test site The additional advantages of using optical fibers include lower cost and immunity from noise‡

12 3 2 Modulators and Detectors

What we have discussed above is just one of the components of a lightwave communication system In addition to this, one requires modulators, which would code the information into the lightwave, and detectors, which could detect the pulses of light at the receiver and decode the information We will discuss briefly the principle behind modulators and detectors

Light sources can either be modulated directly by varying some source input parameter like input current, or the light output can be modulated externally by passing it through devices known as modulators The most promising source to be used in optical fiber communication systems, namely, the semiconductor laser sources, can be modulated easily by varying the input current In fact, in digital systems, the source has to be on–off modulated, and practically, the semiconductor sources can be on–off modulated at high speeds with rise times of less than a nanosecond The on–off modulation is done by biasing the laser diode

† *Laser Focus* September 1979, p 40

‡ *Laser Focus*, October 1977 p 24

slightly below the threshold value, which is typically ~100 mA At this stage, the laser diode operates as an LED and emits incoherent light at a low optical output power An additional current (~20 mA) is added by a high-speed driver, which switches the diode laser from incoherent emission state to a coherent emission with large output optical power By keeping the "off state" slightly below threshold, the delay between the applied electrical pulse and the resulting optical output pulse is minimized, this delay must indeed not be more than the bit interval so that the optical pulses accurately reproduce the input signal

An important factor to be taken care of is the temperature sensitivity of the output optical power In the above-mentioned scheme of operation, this factor can be taken care of by varying the dc bias through an optical feedback circuit so as to take care of both slow changes in ambient temperature and the gradual aging of the laser itself The monitoring of the output power is usually done by collecting the light emanating from the back side of the laser, the light from the front surface being coupled into the fiber itself

For nonsemiconductor laser sources, an external modulator is used for modulation The external modulators make use of various properties possessed by different materials Thus, certain crystals have a birefringence which changes in the presence of an applied electric field Thus, the state of polarization of a beam can be changed by passing it through such a crystal If the crystal is placed between crossed polarizers, one would have an intensity modulation Typical materials used are $LiNbO_3$, $LiTaO_3$, and potassium dihydrogen phosphate (KDP)

Similarly, acousto-optic modulators are based on the interaction of an acoustic beam with the light wave The propagating acoustic wave creates a refractive index grating which in turn diffracts the optical wave

At the receiving terminals or at repeater stations, one requires optical detectors which receive the input optical signal and convert it into electrical signals The three important detector types that find use in lightwave communication are the photomultiplier, the PIN photodiode, and the avalanche photodiode Even though photomultipliers possess large gains, the latter two are expected to find more widespread application because they are less bulky, do not need high bias voltages, and are much cheaper

The simplest solid state photodetector consists of a reverse-biased *p–n* junction with an open center area that is anti-reflection coated to receive the incident light The absorbed photons excite electrons from the valence band into the conduction band The electrons and holes so generated are separated by an applied electric field to induce a photocurrent across the junction

In order to detect very low optical powers one uses the avalanche photodetector. In this device, the electron–hole pair produced by a photon of light is accelerated in the device and is made to release more electron–hole pairs, thus leading to a gain

The photodetectors required in a fiber optic communication system must have a high responsivity at the operating wavelength and must also have sufficient bandwidth in order to accommodate the information rate of the system The most promising photodetector for the 0 8-μm wavelength region seems to be silicon photodiodes They have very fast response times ($\leqslant$ 0 1 nsec) The quantum efficiencies (the ratio of primary photoelectrons generated to the photons incident on the detector) is also large, ~90% For more details on detectors, readers are referred to Melchoir (1977), Personick (1976), Jacobs and Miller (1977)

13

Lasers in Science

13.1. Introduction

In this chapter, we discuss some experiments in physics and chemistry (and related areas) which have been possible only because of the availability of highly coherent and intense laser beams In Sections 13 2, 13 3, and 13 4 we briefly discuss harmonic generation, stimulated Raman emission, and the self-focusing phenomenon, respectively, all these phenomena have opened up new avenues of research after the discovery of the laser In Section 13 5 we outline how laser beams can be used for triggering chemical and photochemical reactions In Sections 13 6, 13 7, and 13 8 we discuss some experiments which can be carried out with extreme precision with the help of lasers Finally, in Section 13 9, we give a fairly detailed account of isotope separation using lasers We would like to point out that there are in fact innumerable experiments that have become possible only with the availability of laser beams, here we discuss only a few of them

13.2. Harmonic Generation

With the availability of high-power laser beams, there has been a considerable amount of work on the nonlinear interaction of optical beams with matter One of the most striking nonlinear effects is the phenomenon of second-harmonic generation, in which one generates an optical beam of frequency 2ω from the interaction of a high-power laser beam (of frequency ω) with a suitable crystal The first demonstration of the second-harmonic generation was made by Franken *et al* (1961) by focusing a 3-kW ruby laser pulse (λ = 6943 Å) on a quartz crystal (see Fig 13 1)

In order to understand the second harmonic generation we consider

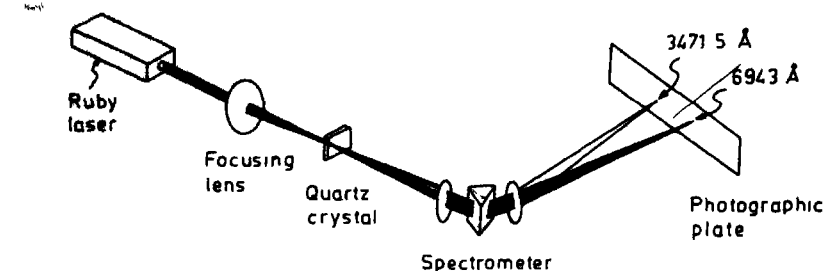



Fig. 13 1 Schematic of the experimental setup used by Franken and his co-workers for the generation and observation of the second harmonic light The beam from the ruby laser is focused by a lens into a quartz crystal which converts a very small portion of the incident light energy (which is at a wavelength of 6943 Å) into its second harmonic which has a wavelength of 3471 5 Å (this wavelength falls in the ultraviolet part of the spectrum) The spectrometer arrangement shown in the figure disperses the components of light at wavelengths 6943 and 3471 5 Å, which can be photographed

the nonlinear dependence of the polarization† on the electric field. It is well known that an atom consists of a positively charged nucleus surrounded by a number of electrons When an electric field is applied, the electron cloud gets displaced and results in a polarization of the crystal For weak intensities of the light beam, the polarization induced in the crystal is proportional to the oscillating electric field associated with the light beam However, for an intense focused beam, the electric fields in the focal region could be very high ($\gtrsim 10^7$ V/cm) and for such strong fields the response of the crystal may be of the type shown in Fig 13 2 From the figure one can see that the electric field in a particular direction can be more effective in polarizing the material than a field in the opposite direction This indeed happens in crystals which have no center of symmetry These crystals are usually piezoelectric in nature‡ In Fig 13 2 we show typical linear and nonlinear dependences of the optical polarization on the electric field which induces the polarization Notice that in the absence of any nonlinearities the curve between the induced polarization and the applied electric field is a straight line, which implies that if the field is increased (or decreased) by a factor α, the polarization would also increase or decrease by the same factor On the other hand, in the presence of nonlinearities, only for weak fields is the curve approximately a straight line

† By "polarization' we imply here the dipole moment induced by the electric field This induced dipole moment is due to the relative displacement of the center of the negative charge from that of the nucleus

‡ A piezoelectric material converts mechanical energy into electrical energy

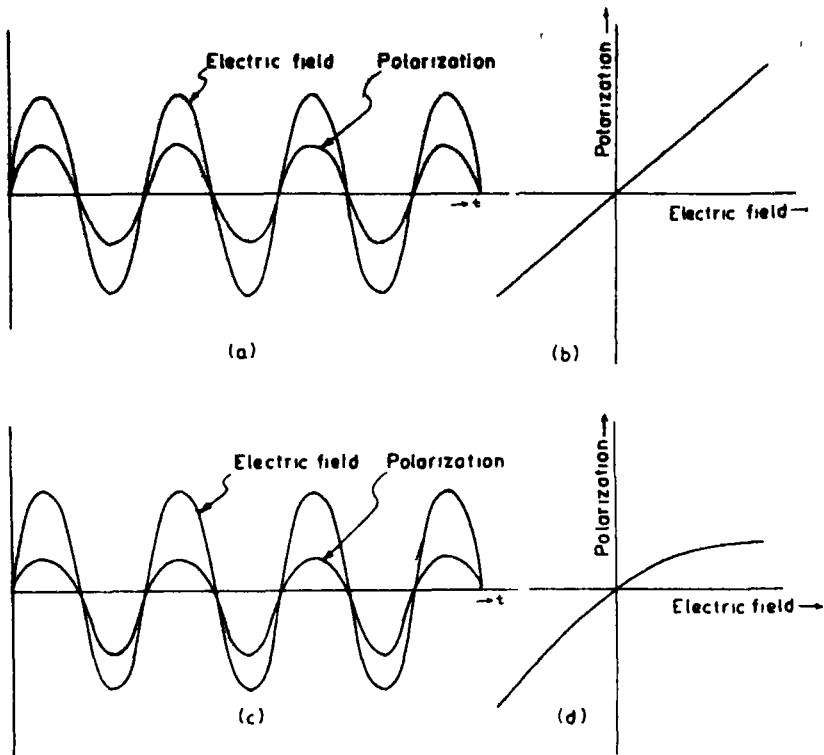



Fig 13 2 The curves show the time variation of the electric field of the light wave incident on the crystal and the induced polarization A typical linear dependence between the induced polarization wave and the applied electric field of the incident light wave is shown in (a) and (c) corresponds to a typical nonlinear dependence The distorted nature of the induced polarization wave in (c) leads to the production of harmonics of the incident light (see Fig 13 3) (b) and (d) show the dependence of the polarization on the applied electric field for a linear and a nonlinear response of a crystal

If we analyze the optical polarization shown in Fig 13 2c, one can show that it approximately consists of three terms (see Fig 13 3)

$$P = P_0 + P_1 \cos \omega t + P_2 \cos 2\omega t \qquad (13\ 2\text{-}1)$$

where ω represents the frequency of the electric field associated with the incident laser beam The presence of the term proportional to $\cos 2\omega t$ results in the second-harmonic generation The term P_0 is known as the dc polarization and corresponds to zero frequency, this dc polarization was also observed by Franken *et al* (1961)

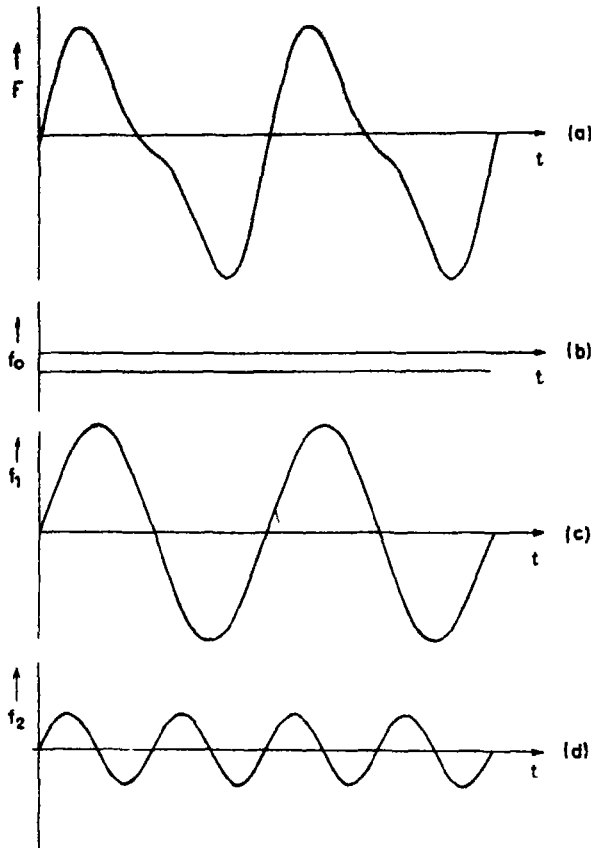



Fig 13 3 By making a Fourier analysis of the wave shown in (a) it can be shown that it consists of three components namely, a zero-frequency component shown in (b), a component with the fundamental frequency ω shown in (c) and a second-harmonic component with a frequency 2ω shown in (d) Thus a nonlinear polarization wave leads to the production of a second harmonic of the fundamental frequency

Figure 13 4† shows how a ruby laser beam (red light) when passed through a crystal of ammonium dihydrogen phosphate gets partially converted into its second harmonic which lies near the ultraviolet region The remaining ruby red light has been filtered by using a glass filter after the crystal

There has been an extensive amount of work in the field of harmonic generation as it provides a coherent beam at frequency 2ω at the output

† Figures 13 4 13 5 and 13 6 will be found following page 278

of the crystal For further details the reader is referred to Baldwin (1969), Zernike and Midwinter (1973), Minck *et al* (1966), Patel (1966), Patel *et al* (1966)

13.3 Stimulated Raman Emission

Another important application of the laser light is in stimulated Raman emission In the Raman effect, a photon of the incident beam is absorbed by a molecule and reemitted at a different frequency If the frequency of the emitted photon is less than that of the incident photon, the emitted line is referred to as the Stokes line and if the frequency of the emitted photon is greater than that of the incident photon, then the emitted line is referred to as the anti-Stokes line The difference in the energies of the incident and the emitted photons corresponds to the vibrational and rotational energy levels of the molecule Thus if ν_e represents the frequency of the emitted photon then

$$\nu_e = \nu_0 \pm \nu_m \qquad (13\ 3\text{-}1)$$

where the positive and the negative signs correspond to the anti-Stokes line and the Stokes line respectively and the frequency ν_m corresponds to rotational or vibrational energy states of the molecule In the stimulated Raman emission, the photons emitted in the ordinary Raman effect are made to stimulate further Raman emissions A layout of an experimental arrangement for observing stimulated Raman emission from Raman laser material (e g , nitrobenzene) is shown in Fig 13 5 An intense laser beam is focused on the Raman laser material The photons corresponding to the Stokes lines have a wavelength in the infrared region which does not appear on the color film In the anti-Stokes emission, the frequency of the emitted photon is higher, which results in orange, yellow, and green rings (see Fig 13 6)†

The stimulated Raman effect has been observed in a large number of materials, which provides us with hundreds of coherent sources of light from ultraviolet to infrared

Stimulated Raman scattering and continuous wave Raman oscillation have also been observed in optical fiber waveguides‡ by many workers using low pump powers (e g , Stolen *et al* , 1972, Lin *et al* , 1977a Lin *et*

† That the emitted photons of a particular frequency should appear in well-defined cones about the direction of the incident photons follows from the conservation of momentum The direction of the emitted photon can be found by using the fact that a photon of frequency ν has a momentum equal to $h\nu/c$

‡ Optical fiber waveguides are discussed in Chapter 12, for a detailed discussion on optical fibers see e g , Ghatak and Thyagarajan (1980) and Gloge (1976)

al, 1977b). Raman oscillation in optical fiber waveguides requires low pumping powers because of the strong transverse confinement of the optical energy in the waveguide and the very long interaction lengths (~600 m) that are possible with the use of optical fibers The pumping source is usually the 1 064-μm radiation of the Nd YAG laser (see Section 9 5) or the Ar-ion laser The broad Raman gain bandwidth of silica glass fibers has indeed been used in making a tunable continuous wave fiber Raman oscillator (Stolen *et al*, 1977, Lin *et al*, 1977a) Such Raman oscillators can indeed be used as tunable light sources and are especially useful for fiber optical communication studies as the region of operation at 1 3 μm wavelength is indeed becoming very important due to the existence of a region of zero dispersion in single-mode glass fibers

13.4. Self-Focusing

Another interesting nonlinear effect is the self-focusing phenomenon This arises because of the dependence of the refractive index of the material on the intensity of the beam, this dependence can be written in the following form

$$n = n_0 + n_2 I \quad (13\ 4\text{-}1)$$

where n_0 and n_2 are constants and I represents the intensity of the laser beam Thus, if the intensity† of the beam is maximum on the axis and decreases radially, and if n_2 is positive, then the beam would get focused (see Fig 13 7) This can be qualitatively understood from the fact that the velocity (which is inversely proportional to the refractive index) will be

† This is indeed the case for a laser beam where one usually has a Gaussian variation of intensity along the wave front

Fig 13 4 When a ruby laser beam (red light) is passed through a crystal of potassium dihydrogen phosphate a portion of the energy gets converted into blue light, which is the second harmonic The remaining ruby red light is suppressed by using a glass filter after the crystal The two beams are made visible by means of smoke-filled glass troughs (Photograph courtesy Dr R W Terhune)

Fig 13 5 Schematic of the experimental setup used for observation of stimulated Raman emission The light emerging from the ruby laser is focused by a lens into a cell containing benzene The Stokes lines occur in the infrared and the anti-Stokes lines occur in the visible region The various anti-Stokes lines appearing in the red orange yellow and green regions of the spectrum form rings around the central ruby laser spot (see Fig 13 6)

Fig 13 6 The stimulated Raman effect obtained by focusing a ruby laser beam in a cell filled with benzene and photographing the scattered radiation in a setup similar to that shown in Fig 13 5 (Photograph courtesy Dr R W Terhune)

FIGURE 13 4

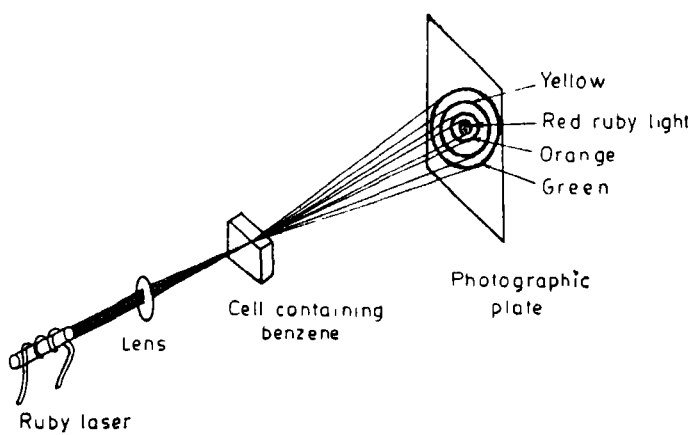



FIGURE 13 5

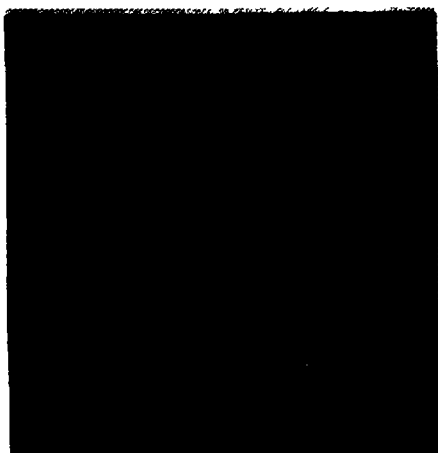



FIGURE 13 6

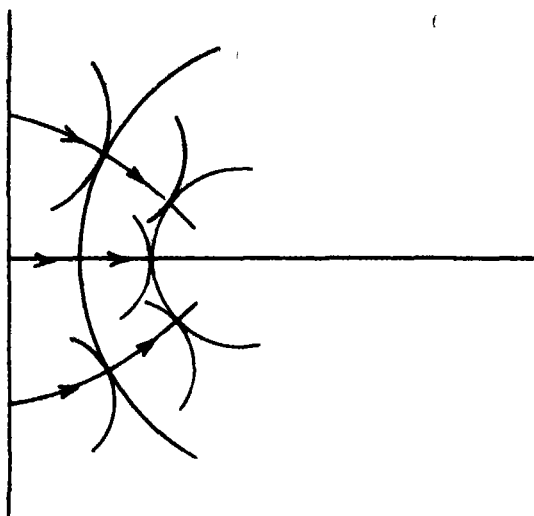



Fig 13 7 Self-focusing of a beam occurs when a high-power laser beam with an intensity distribution which decreases away from the center passes through a medium with a positive value of n_2

minimum on the axis, and by simple Huygens' construction one can show that a plane wave front would become converging This results in what is known as the self-focusing phenomenon For further details the reader is referred to Sodha, Ghatak, and Tripathi (1974, 1976), Svelto (1975), and Khokhlov *et al* (1976)

13 5 Lasers in Chemistry

Lasers are expected to find important applications in chemistry Because of the extremely large temperatures obtainable at the focus of a laser beam, the laser is an excellent tool for triggering chemical and photochemical reactions† Electric fields larger than 10^9 V/cm are obtainable at the focus, such fields are larger than the fields that hold the valence electrons to the atoms

Lasers attached to microscopes are used as microprobes in microanalysis Such microanalysis can give information regarding the presence of trace metals in various tissues The technique essentially involves sending a giant pulse to a preselected area and vaporizing some of the

† The use of lasers in photophysics and photochemistry has been discussed at a popular level by Letokhov (1977)

target material The vapor so produced may itself emit a spectrum of wavelengths or one may pass a discharge through it to produce a characteristic spectrum An analysis of the spectrum gives the presence of various elements

It has been demonstrated that molecules that have been excited by an infrared laser react faster than the molecules that are in the ground state The extremely high monochromaticity of the laser allows one to selectively excite different bands of a molecule and thus leads to the possibility of producing some new chemical products

With the generation of intense laser pulses lasting for a few picoseconds†, one can now study ultrafast physical and chemical processes This gives an opportunity for understanding the most fundamental processes with unprecedented time resolution The technique is essentially to excite the sample with an intense pulse and then to study the behavior of a certain characteristic parameter of the sample (e g , absorption or scattering) as a function of the delay time after excitation Such studies have indeed been successfully applied in studying various processes such as the redistribution of light energy absorbed by cholorophyll in the photosynthetic process, to observe ultrafast chemical reactions, to obtain the vibrational and rotational decay constants of molecules, to study photovisual processes, and others For a review of some of these applications the reader is referred to the articles by Alfano and Shapiro (1975) and Busch and Rentzepis (1976)

13 6. Lasers and Ether Drift

Lasers have made possible an experiment to test the presence of ether drift with an accuracy a thousand times better than could be obtained before A setup of the experiment is shown in Fig 13 8 The beams from two lasers oscillating at slightly different frequencies are combined with a beam splitter and detected by a photomultiplier The slight difference in frequency between the two lasers produces beats at a frequency equal to the difference between the two frequencies The oscillation frequency of the laser depends on the length of the resonant cavity and also on the speed of light in the cavity Thus a rotation of the apparatus must change the frequency of oscillation of the lasers and hence the beat frequency, if there was any ether drift. When the experiment was performed no change in beat frequency was observed The apparatus was sensitive enough to detect a velocity change as small as 0 03 mm/sec

† 1 psec = 10^{-12} sec

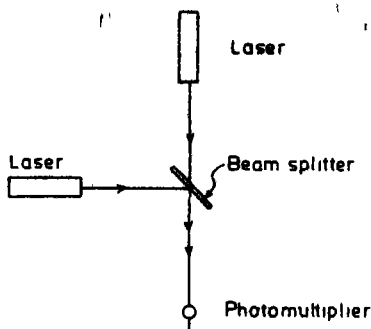



Fig 13 8 An experimental setup for detecting the presence of ether drift

13 7. Rotation of the Earth

Lasers have been used to detect the absolute rotation of the Earth If light is made to rotate in both clockwise and anticlockwise directions around a square with the help of mirrors as shown in Fig 13 9, then if the square is at rest, the time taken for light to travel around the square in both the clockwise and the anticlockwise directions would be the same

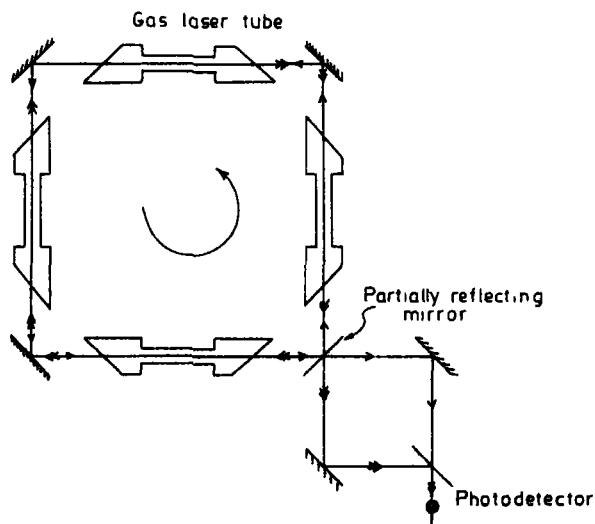



Fig 13 9 A ring laser for detecting the absolute rotation of the earth When the system is at rest, both the clockwise and anticlockwise rotating beams have the same frequency When the system is rotated about an axis normal to the plane containing the system, the clockwise rotating and the anticlockwise rotating beams have slightly different frequencies When they are combined, then beats are produced which can easily be detected

But if the square is rotated about an axis which is normal to the plane of the square, then the time taken for light to travel along one direction will be different from that taken along the other direction Thus if one could measure this difference one could obtain information about the rotation of the square†

An experiment to detect the rotation of the Earth by using such a method with ordinary light sources was performed by Sagnac in 1914 and then by Michelson and Gale in 1925 With the use of lasers, one can do similar experiments with much more precision and sensitivity The sides of the square are gas discharge tubes containing helium and neon The corners of the square are occupied by mirrors (see Fig 13 9) Light beams traveling along either direction would undergo amplification Thus one would have two beams, one propagating in the clockwise direction and the other in the anticlockwise direction The frequency of oscillation of the laser would depend on the path length along the square Since the path lengths along the two directions are different when the system is rotating in the plane, two different frequencies are obtained By mixing the light beams of the two frequencies, one can detect the beat frequency of the beams and hence the rate of rotation of the square For example, at New York (which is at a latitude of 40°40′N) the effective speed of rotation is about 1/6 of a degree per minute, this would correspond to a beat frequency of 40 Hz

In the ring interferometer that we have discussed, one can show that the phase difference introduced by the counterpropagating beams is (see, e g , Post, 1967)

$$\Delta\varphi = \frac{8\pi}{c\lambda}\mathbf{\Omega}\cdot\mathbf{A} \qquad (13\ 7\text{-}1)$$

where $\mathbf{\Omega}$ is the rotation vector and $\mathbf{A}$ is the area enclosed by the optical path This phase change is generally too small for direct measurement in the range of rotation speeds encountered in inertial navigation For example, a rotation rate of 5×10^{-7} rad/sec (0 1 deg/h) over an area of 0 1 m^2 at $\lambda = 0\ 6\ \mu$m yields a phase shift of 7×10^{-9} rad ($\sim 10^{-9}$ of a fringe!) The availability of extremely low loss single-mode optical fibers makes it possible to increase $\Delta\varphi$ by 3 to 4 orders of magnitude by increasing the effective area A by having the light beams propagate through a large number of turns of an optical fiber In addition to this increase in area, it is possible to have both the clockwise and counterclockwise beams follow identical paths, thus stabilizing the differential path lengths

Extensive research programs are underway at present in many

† The difference in path length between the two paths is extremely small, thus only a shift of a hundred-thousandth of a wavelength would be produced when the square is of side 3 m and is kept at a latitude of 40 degrees on the surface of the earth

laboratories around the world to develop a fiber optic rotation sensor for inertial navigation purposes Work is also in progress for using integrated optical circuits for signal processing in the sensors, this should lead to mass production of completely guided reciprocal interferometers (Arditty *et al*, 1980, Vali and Shorthill, 1977, Goldstein and Goss, 1979)

13 8. Photon Statistics

Let us consider an experiment in which a beam of light from a source is allowed to fall on a phototube for a specific time interval T by having a shutter open in front of the detector for the time T, one then registers the number of photoelectrons so liberated Then the shutter is again opened for an equal time interval T after a certain time delay which is longer than the coherence time of the source and the number of photoelectrons counted during the interval is again registered This experiment is repeated a large number of times ($\sim 10^5$) and the number of photoelectrons produced in equal intervals of time is counted The results of such an experiment give one the probability distribution $p(n, T)$ of counting n photons in a time T Here it is assumed that the light source is stationary, i e , the long time average of the intensity is fixed and independent of the particular long time period chosen for measuring it The above-obtained probability distribution contains information regarding the statistical properties of the source, and such studies have applications in spectroscopy, stellar interferometry, etc

It can be shown that for a polarized thermal source, for counting times T much smaller than the coherence time T_c, the photoelectron counting distribution $p(n)$ is given by† the following formula (see, e g ,

† It is much beyond the scope of this book to go into the derivation of Eqs (13 8-1) and (13 8-2) However it may be worthwhile to mention that the probability distribution $p(n, t\ T)$ of registering n photoelectrons by an ideal detector in a time interval t to $t + T$ is given by (see, e g , Mandel, 1958, 1959)

$$p(n, t, T) = \int_0^\infty \frac{(\alpha w)^n}{n!} e^{-\alpha w} P(w)\, dw$$

where α is the quantum efficiency of the detector and $w[= \int_t^{t+T} I(t')\, dt']$ is the integrated light intensity and $P(w)$ is the probability distribution corresponding to the variable w Thus the photoelectron counting distribution given by the above equation depends on the particular form of the probability distribution $P(w)$ Now, for a polarized thermal source, it can be shown that

$$\begin{aligned} P(w) &= (1/\langle w\rangle)e^{-w/\langle w\rangle} && \text{for } T \ll T_c \\ &= \delta(w - \langle w\rangle) && \text{for } T \gg T_c \end{aligned}$$

where $\langle w\rangle = T\langle I\rangle$ and angular brackets denote averaging Further, for an ideal laser, the beam would not exhibit any intensity fluctuations and $P(w) = \delta(w - \langle w\rangle)$ On substituting the above equations for $P(w)$ in the equation for $p(n, t, T)$, one gets Eqs (13 8-1) and (13 8-2)

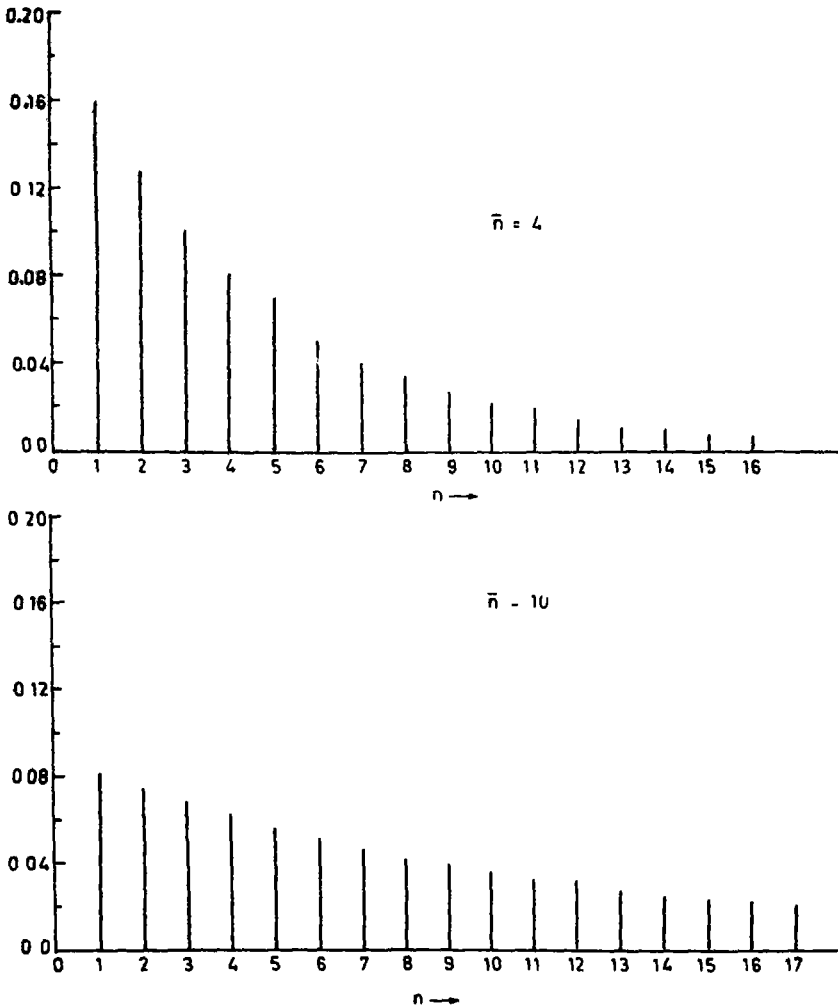



Fig 13 10a The probability distribution for the arrival of photons from a thermal source

Loudon, 1973)

$$p(n) = \frac{\langle n \rangle^n}{[1 + \langle n \rangle]^{n+1}} \tag{13 8-1}$$

where $\langle n \rangle$ represents the mean number of counts in time T Figure 13 10a shows typical plots of the above distribution for $\langle n \rangle = 4$ and $\langle n \rangle = 10$

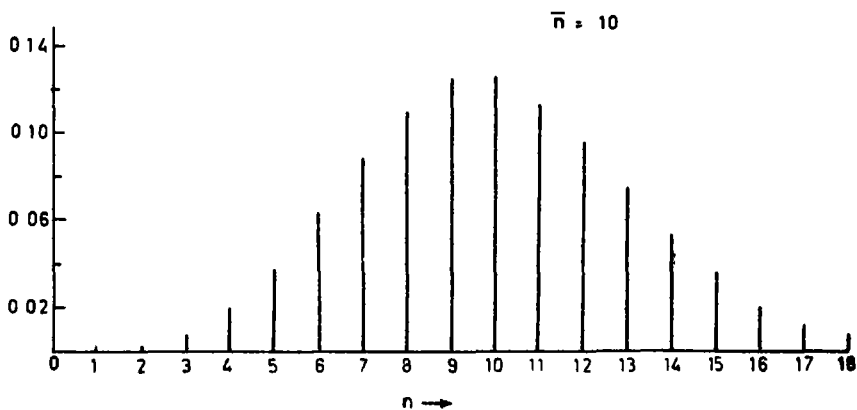
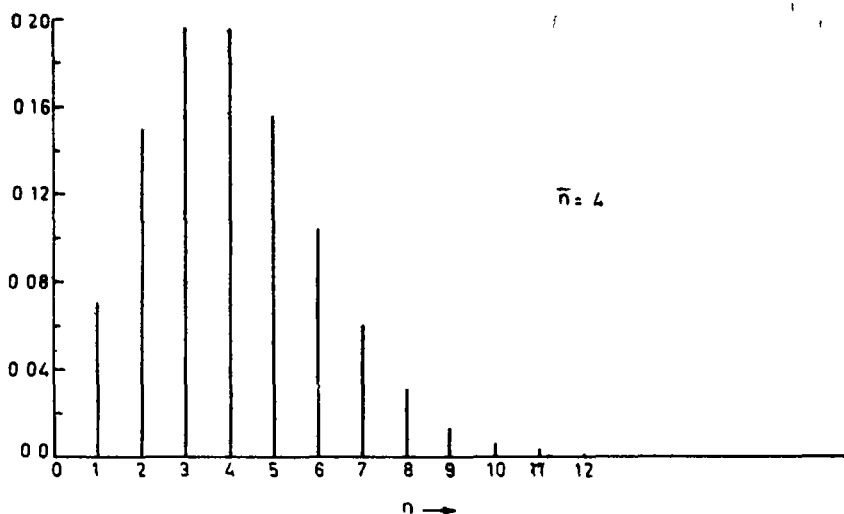



Fig 13 10b The probability distribution for the arrival of photons from a laser source corresponds to the Poisson distribution

On the other hand, for very large values of T as compared to T_c, all the fluctuations in the intensity may be expected to be averaged out during the counting period For such a case, the photon counting distribution is given by

$$p(n) = \frac{\langle n \rangle^n}{n!} e^{-\langle n \rangle} \qquad (13\ 8\text{-}2)$$

which is a Poisson distribution

When the source is an ideal laser, the beam would not exhibit any intensity fluctuations and the counting distribution would be again given by Eq (13 8-2)—see also Eq (8 4-6)

Figure 13 10b shows the probability distribution given by Eq (13 8-2) for $\langle n \rangle = 4$ and $\langle n \rangle = 10$ As can be observed from the figure, even for a beam of constant intensity the fluctuations still exist

Experiments on photoelectron counting were first carried out by Arecchi *et al* (1966a,b), in which they have shown that for a laser operating much beyond threshold, the counting distribution is indeed Poissonian The results on the counting distribution from a laser source near threshold, from thermal sources and from a mixture of a laser and a thermal source, have also been discussed by Arecchi *et al* (1966a,b) For further details, the reader is referred to Arecchi (1976), Mandel and Wolf (1970), and Mehta (1970)

13.9. Lasers in Isotope Separation†

A new application of lasers (in particular, tunable lasers), that has been much discussed in recent years, is in isotope separation The technique seems to promise efficient separation processes that may perhaps revolutionize the economics of the separation and the use of isotopes The major interest in the so-called LIS (laser isotope separation) process would be its possibility for large-scale enrichment of uranium for use in nuclear power reactors There are, however, other applications for pure isotopes in medicine, science, and technology, if they could be produced economically We shall, in the following, briefly discuss the principle of LIS and the variety of options which have been considered and already demonstrated on the laboratory scale

Isotopes are atoms that have the same number of protons and electrons but which differ in the number of neutrons Since most chemical properties are determined by the electrons surrounding the nucleus, the isotopes of an element behave in almost indistinguishable ways Thus one of the common methods of separating an element from a mixture (by making use of its chemical properties) becomes cumbersome when an isotope is to be separated from a mixture with another isotope of the same element Light elements lying below oxygen in the Periodic Table may be separated using repeated chemical extraction Isotopes of heavier atoms may be separated using physical methods For example, because of

† The material in this section was kindly contributed by Dr S V Lawande of Bhabha Atomic Research Centre Bombay

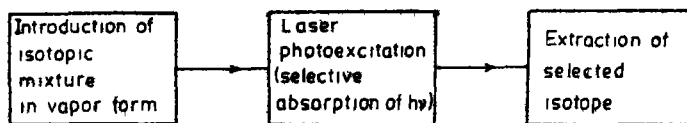



Fig 13 11 Block diagram of a typical laser isotope separation process A laser excites one of the isotopes from the isotopic mixture through selective absorption and the excited isotope atoms are separated using one of the many techniques

the differences in masses of the isotopes, they diffuse at different rates through a porous barrier and repeated passage through various stages leads to the required concentration of the isotope to be separated

Isotope separation using a laser beam is a fundamentally different technique where one makes use of the slight differences in the energy levels of the atoms of the isotopes due to the difference in nuclear mass This difference is termed the isotope shift Thus, light of a certain wavelength may be absorbed by one isotope while the other isotope of the element may not absorb it† Since the light emerging from a laser is extremely monochromatic, one may shine laser light on a mixture of two isotopes and excite the atoms of only one of the isotopes, thus earmarking it for subsequent separation A block diagram of a typical LIS process is shown in Fig 13 11 It may be of interest to note here that the basic physical idea of isotope separation by light was conceived more than 50 years ago The first successful separation of mercury isotope (^{202}Hg) was reported by Zuber (1935), who irradiated a cell containing natural mercury vapor with the light from a mercury lamp Since the invention of tunable lasers during the last few years, it has become possible to revive the interest in photochemical separation processes for large-scale isotope separation

In addition to the high monochromaticity, the high intensity of the laser is also responsible for its application for isotope separation because with low-intensity beams the separation rate would be too little for practical use

The basic principle behind the laser isotope separation process is to first selectively excite the atoms of the isotope by irradiating a stream of the atoms by a laser beam and then separate the excited atoms from the mixture Various techniques exist for separation We will discuss a few of them, for more details the reader is referred to Zare (1977)

† When an atom absorbs light, it jumps from one energy level to another, the difference in energy between the two levels being just equal to the energy of the incident photon Since the energy levels are slightly different for two different isotopes, their absorption properties are also different The isotope shift between hydrogen and deuterium is given by $\Delta\nu/\nu \simeq 2\,7 \times 10^{-4}$ For uranium, the isotope shift is given by $\Delta\nu/\nu \simeq 0\,6 \times 10^{-4}$

13 9 1 Separation Using Radiation Pressure

An interesting method of laser isotope separation is the deflection of free atoms or molecules by radiation pressure A photon of energy $h\nu$ carries with it a momentum of $h\nu/c$ When this photon is absorbed by an atom, conservation of momentum requires that the atom acquire this momentum Thus the absorption tends to push the atom in the direction of travel of the incident photon The momentum acquired in a single absorption is very small, hence for the atom to gain sufficient momentum, it must absorb many photons This requires that the atoms have a short lifetime in the excited state before dropping back to the ground state It should be noted that every time an atom emits a photon, it acquires a momentum equal and opposite to that it gained during absorption Since the emissions occur in all random directions, the net effect of many absorptions and emissions is to push the atoms along the laser beam In the present technique, a laser beam is allowed to impinge on an atomic beam at right angles (see Fig 13 12) and the atoms of the isotope (which absorb the radiation) are deflected by the laser beam

To give some idea of the numbers involved, the velocity resulting from the momentum transfer is about 3 cm/sec in the case of sodium atoms After an excitation event the atom/molecule remains in the excited state for a certain time $\tau \sim 10^{-8}$ sec Thus a sodium atom traveling with a thermal velocity $v \approx 10^5$ cm/sec through an interaction zone of length 1 cm will be subjected to 10^3 absorption events and gain a net velocity of 3×10^3 cm/sec in the direction of the laser beam The resulting deflection of 30 mrad will be sufficient for separating the sodium atom Such a scheme has been used to separate the isotopes of barium (Bernhardt *et al*, 1974) More sophisticated modifications of this scheme which improve the photon economy have also been conceived

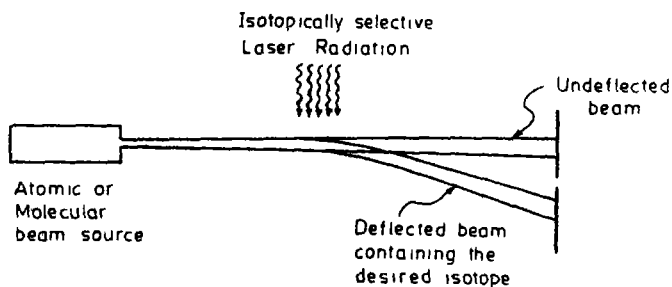



Fig 13 12 Separation of isotope by deflection caused by selective absorption The atomic beam emerging from the source is impinged by a laser beam tuned to excite atoms of the isotope to be separated The absorption causes the atoms to acquire a momentum and by repeated absorption they gain enough kinetic energy to get deflected from the main beam

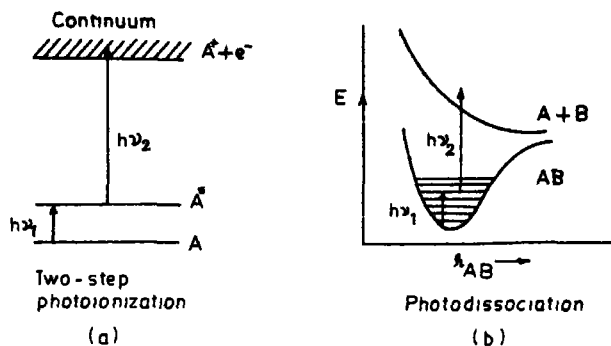



Fig 13 13 (a) Two-step photoionization of atoms, (b) two-step photodissociation of molecules The ionized atoms can be separated by application of an electric field The dissociation products from a molecule may be separated by chemical reactions

13 9 2 Separation by Selective Photoionization or Photodissociation

The most popular and perhaps universally applicable scheme of isotope separation is the two-step photoionization of atoms or the two-step dissociation of molecules (Fig 13 13) The first step causes the selective excitation, this is followed by a second excitation which ionizes the excited atoms or dissociates the excited molecules In the case of atoms the separation can be carried out by extracting the ions by means of electric fields In the case of molecules the dissociation products must be separated from the other molecules This may be carried out directly or by means of chemical reactions It must be mentioned here that the two-step photoionization was used to demonstrate the feasibility of LIS for uranium at the Lawrence Livermore Laboratory in the USA† In this experiment an atomic beam of uranium, generated in a furnace at a temperature of about 2100°C, was excited by the light of a dye laser (isotope-selective excitation) and then ionized by the light of a high-pressure mercury lamp and the ^{235}U isotope which is present in natural uranium in a concentration of 0 71% was enriched to 60%

The application of two-step excitation to molecules is described in an experiment on the separation of the isotopes† ^{10}B and ^{11}B In this experiment $^{11}BC_3$ isotope was selectively excited by the light of a CO_2 laser which emits lines corresponding to vibration transitions of $^{11}BC_3$ The molecules excited in this manner were dissociated by light with a wavelength between 2130 and 2150 Å. The fragments generated by this dissociation, originating mainly from $^{11}BC_3$, were bound by reaction with

† Reported in *Physics Today* **27**(9), 17 (1974)

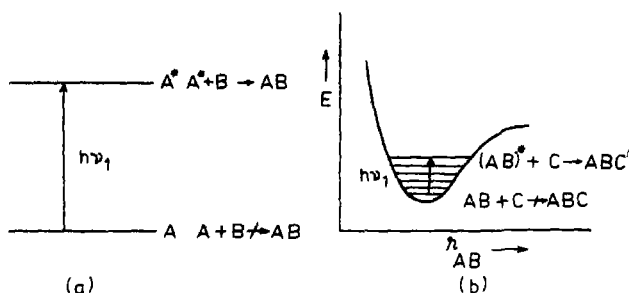
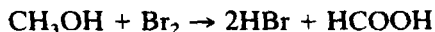



Fig 13 14 Chemical reactions that take place only with excited atoms or molecules may be employed for the separation of the excited isotope atoms or molecules

O_2 It was found that with five light pulses of the CO_2 laser radiation, a 14% isotopic enrichment of a 5-μg sample could be obtained

13 9 3 Photochemical Separation

Another possible way of separating selectively excited atoms/molecules from those in the ground state is by means of a chemical reaction The reaction must be so chosen that it takes place only with atoms or molecules in the excited state but not with those in the ground state The isotope of interest can be separated from the other isotopes present by the chemical separation of the reaction products The basic idea is illustrated in Fig 13 14 An example of a separation using this scheme is the enrichment of deuterium using an HF laser [Mayer *et al*, 1970)] Some lines of the HF laser coincide with strong transitions of methanol but not with the corresponding lines of deuteromethanol The excitation activates the reaction

$$CH_3OH + Br_2 \rightarrow 2HBr + HCOOH$$

A one-to-one gas mixture of $CH_3OH + CD_3OD$ can be converted under irradiation for 60 sec with a 90-W HF laser, in the presence of Br_2 into a mixture containing 95% CD_3OD

One of the most important fields in which the laser isotope separation process would find application is in the nuclear power industry, which requires uranium enriched with the isotope of mass number 235 The present method of enrichment is through gaseous diffusion through a number of stages This process is quite costly, the cost of obtaining uranium 235 of 90% purity is about 2 3 cents per milligram (Zare, 1977) Similarly, the cost of other isotopes of such a concentration is also high In addition to the nuclear power industry, the isotope separation process would also help obtain isotopes used as tracers in medicine, agriculture, research, industry, etc

14

Lasers in Industry

14.1 Introduction

In Chapter 9 we discussed the special properties possessed by laser light, namely, its extreme directionality, its extreme monochromaticity, and the large intensity associated with some laser systems In the present chapter, we briefly discuss the various industrial applications of the laser

The beam which is coming out of a laser is usually a few millimeters (or more) in diameter and hence, for most material processing applications, one must use focusing elements (like lenses) to increase the intensity of the beam The beam from a laser has a well-defined wave front, which is either plane or spherical When such a beam passes through a lens, then according to geometrical optics, the beam should get focused to a point In actual practice, however, diffraction effects have to be taken into consideration, and one can show that if λ is the wavelength of the laser light, a is the radius of the beam, and f is the focal length of the lens, then the incoming beam will get focused into a region of radius (see Fig 14 1)†

$$b \simeq \lambda f/a \qquad (14\ 1\text{-}1)$$

As can be seen, the dimension of this region‡ is directly proportional to f and λ (the smaller the value of λ, the smaller the size of the focused spot) and inversely proportional to the radius a If P represents the power of the laser beam, then the intensity I, obtained at the focused region, would be given by

$$I \simeq \frac{P}{\pi b^2} \simeq \frac{Pa^2}{\pi \lambda^2 f^2} \qquad (14\ 1\text{-}2)$$

† Here we have assumed that the aperture of the lens is greater than the width of the beam If the converse is true then a would represent the radius of the aperture of the lens

‡ The dimension of the focused region is usually larger than that given by Eq (14 1-1) due to the multimode emission of the laser We are also assuming here that the lenses are aberrationless In general aberrations increase the spot dimension resulting in lower intensities

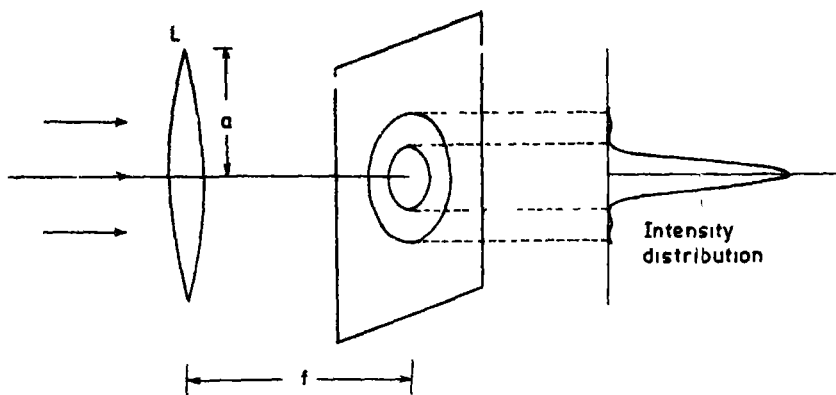



Fig 14 1 When a plane wave of wavelength λ falls on a lens of radius a then at the focal plane F of the lens one obtains an intensity distribution of the type shown in the figure About 84% of the total energy is confined within a region of radius $\lambda f/a$

Thus if we focus a 1-W laser beam (with λ = 1 06 μm and having a beam radius of about 1 cm)† by a lens of focal length 2 cm, then the intensity obtained at the focused spot would be given by

$$I \approx \frac{1}{3\ 14 \times (1\ 06 \times 10^{-4} \times 2)^2} \text{ W/cm}^2$$

$$\approx 7 \times 10^6 \text{ W/cm}^2 \qquad (14\ 1\text{-}3)$$

Figure 14 2 shows the spark created in air at the focus of a 3-MW peak power giant pulsed ruby laser The electric field strengths produced at the focus are of the order of 10^7 V/cm Notice here that such large intensities are produced in an extremely small region whose radius is $\sim 2 \times 10^{-4}$ cm Further, as can be seen from Eq (14 1-2), the larger the value of a, the greater will be the intensity, as such, one often uses a beam expander to increase the diameter of the beam, a beam expander usually consists of a set of two convex lenses as shown in Fig 14 3

It may be noted that when one produces such small focused laser spots, the beam has a large divergence, and hence near the focused region, the beam expands again within a very short distance This distance (which may be defined as the distance over which the intensity of the beam drops to some percentage of that at the focus) defines the depth of focus Thus, smaller focused spots lead to a smaller depth of focus This must also be kept in mind while choosing the parameters in a laser processing application

† The 1·06-μm radiation is emitted from the neodymium-doped YAG or glass laser (see Section 9 5)

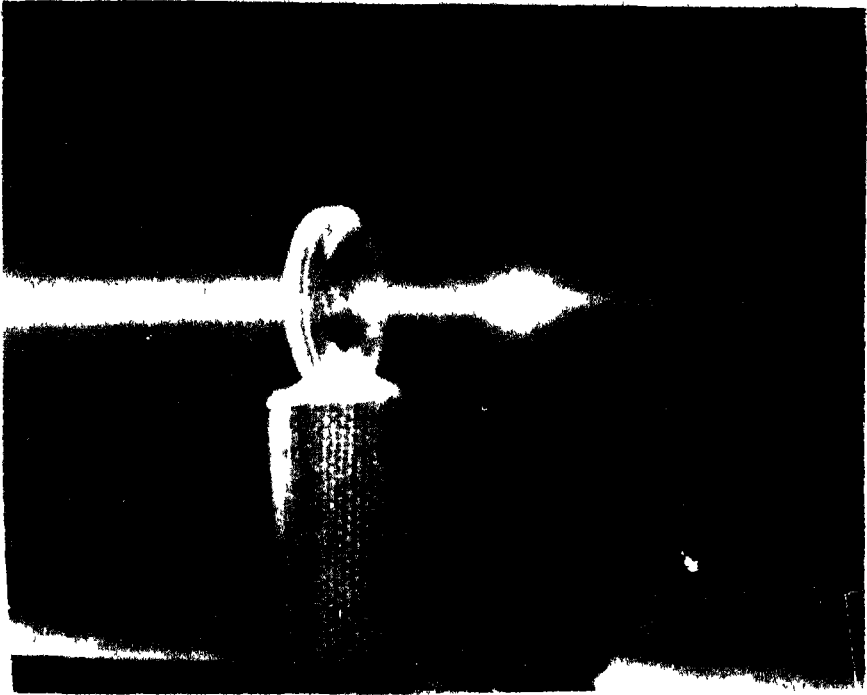



Fig 14 2 Spark created in air at the focus of a 3-MW peak power giant pulsed ruby laser The electric field strengths produced at the focus are of the order of 10^7 V/cm (Photograph courtesy Dr R W Terhune)

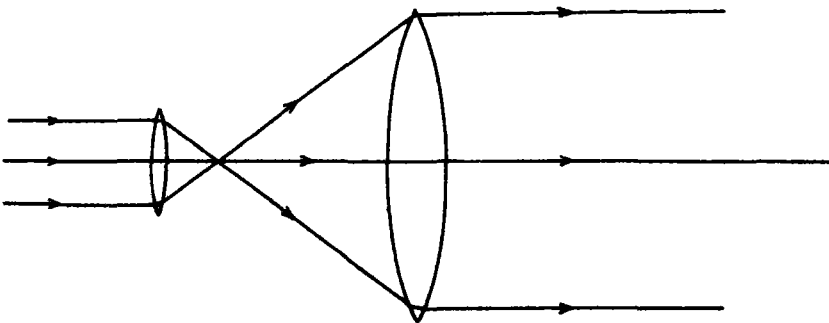



Fig 14 3 A beam expander consisting of two convex lenses

We now discuss in the next few sections some of the important applications of the laser in industry

14.2. Applications in Material Processing

14.2.1. Laser Welding

High-power lasers have found important applications in welding For example, a weld of 1/4-in -thick stainless steel was carried out by a CO_2 laser having an output power of 3 5 kW The material was moved at a speed of 2 cm/sec in the focal plane of a lens of focal length 25 cm

Pulsed ruby lasers have also been used in welding For example, a pulsed ruby laser beam having an energy of 5 J with pulse length of about 5 nsec was used in welding 0 18-mm-thick stainless steel The weld was made using overlapping spots and the laser was pulsed at a rate of 20 pulses per minute The focused spot was about 1 mm in diameter and the associated power density was $\sim 6 \times 10^5$ W/cm^2, the data are from Charschan (1972)

Laser welding has found important applications in the fields of electronics and microelectronics which require precise welding of very thin wires (as small as 10 μm) or welding of two thin films together In this field, the laser offers some unique advantages Thus, because of the extremely short times associated with the laser welding process, welding can be done in regions adjacent to heat-sensitive areas without affecting these elements Figure 14 4 shows a weld performed with a laser on a transistor unit Further, welding in otherwise inaccessible areas (like inside a glass envelope) can also be done using a laser beam Figure 14 5 shows such an example in which a 0 03-in wire was welded to a 0 01-in -thick steel tab without breaking the vacuum seal In laser welding of two wires, one may have an effective weld even without the removal of the insulation

Laser welds can easily be performed between two dissimilar metals Thus, a thermocouple may easily be welded to a substrate without much damage to adjacent material One can indeed simultaneously form the junction and attach the junction to the substrate This method has been used in attaching measuring probes to transistors, turbine blades, etc Laser weld not only achieves welding between dissimilar metals but also allows precise location of the weld

In welding, material is added to join the two components Thus the laser power must not be too high to evaporate the material, removal of material leads, in general, to bad welds Thus the laser used in welding processes must have a high average power rather than high peak power

Fig 14.4 The arrow shows the position of the weld per[illegible] with a laser on a transistor unit (After Gagliano *et al* 1969)

The neodymium–YAG lasers and carbon dioxide lasers are two important kinds of lasers that find wide-ranging applications in welding

14.2.2 Hole Drilling

Drilling of holes in various substances is another interesting application of the laser† For example, a laser pulse having a pulse width of about 1/1000 of a second and an energy of approximately 0.05 J can burn

† In the early 1960s, the power of a focused laser beam was measured by the number of razor blades that the beam could burn through simultaneously the "gillette" being the unit of measurement of power per blade burnt through

Fig. [illegible] Laser welding in inaccessible areas. The photograph shows the welding of a 0.003-in. wire to a 0.01-in. thick steel tab inside a vacuum tube without breaking the vacuum seal. The arrows point towards the repaired connections. (After Weaver, 1971. Photograph courtesy Dr. Weaver.)

through a 1-mm-thick steel plate leaving behind a hole of radius ~0.1 mm Further, one can use a laser beam for the drilling of diamond dies used for drawing wires. Drilling holes less than about 250 μm in diameter by using metal bits becomes very difficult and is also accompanied by frequent breakage of drill bits With lasers one can easily drill holes as small as 10 μm through the hardest of substances The Swiss watch industry in Europe has been using flash-pumped neodymium–YAG lasers to drill ruby stones used in timepieces. In addition to the absence of problems like drill breakage, laser hole drilling has the advantage of precise location of the hole For further details the reader is referred to Charschan (1972)

14 2 3 Laser Cutting

Lasers also find application in cutting materials The most common laser that is used in cutting processes is the carbon dioxide laser due to its high output power

In the cutting process, one essentially removes the materials along the cut When cuts are obtained using pulsed lasers, then the repetition frequency of the pulse and the motion of the laser across the material is adjusted so that a series of partially overlapping holes are produced The width of the cut should be as small as possible with due allowance to avoid any rewelding of the cut material The efficiency of laser cutting can be increased by making use of a gas jet coaxial with the laser (see Fig. 14 6)

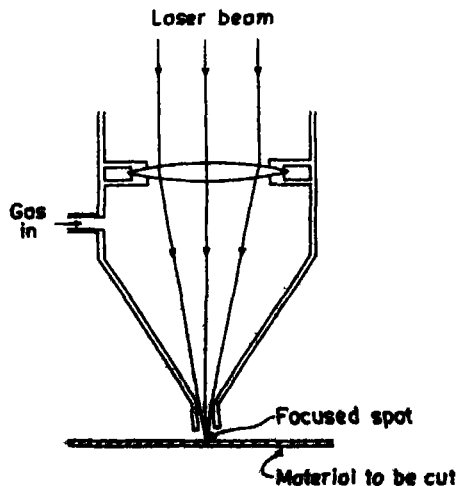



Fig. 14 6 A gas-jet-assisted laser cutting scheme The gas jet assists in expelling the molten material from the cut

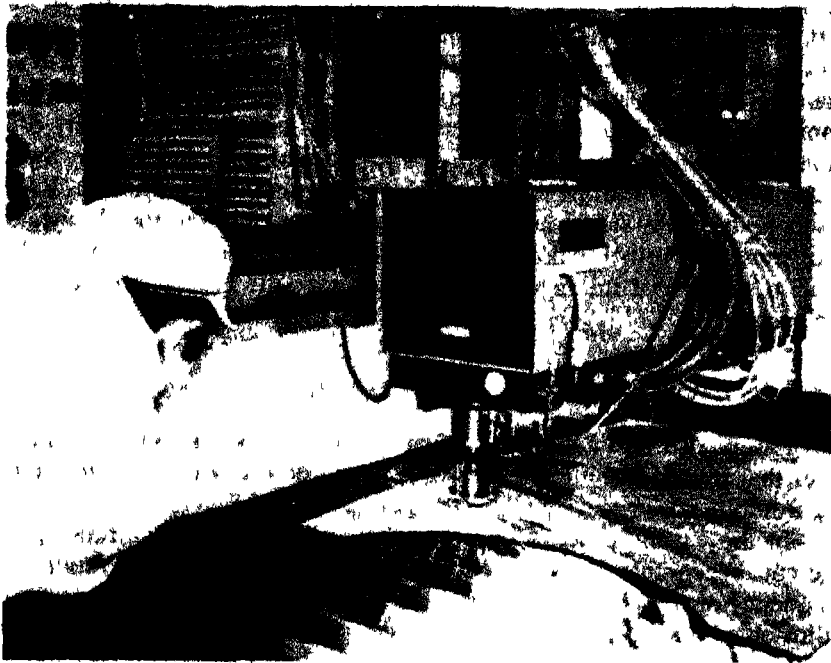



Fig 14 7 Cutting wood with a gas-jet-assisted carbon dioxide laser (Photograph courtesy Ferranti Ltd)

In some cases one uses a highly reactive gas like oxygen so that when the laser heats up the material, it interacts with the gas and gets burnt The gas jet also helps in expelling the molten material Such a method has been used to cut materials like stainless steel, low-carbon steel, titanium etc For example, a 0 13-cm-thick stainless steel plate was cut at the rate of 0 8 m/min using a 190-W carbon dioxide laser using oxygen jet

In some methods, one uses inert gases (like nitrogen or argon) in place of oxygen Such a gas jet helps in expelling the molten materials Such a technique would be very efficient with materials which absorb most radiation at the laser wavelength Wood, paper, plastic, etc have been cut using such a method A gas-jet-assisted CO_2 laser can be used for obtaining parallel cuts of up to 50 mm depth in wood products At the cut edges carbonization occurs, but it is usually limited to a small depth (about a few tens of μm) of the material This causes a discoloration only and can be decreased by increasing the cutting speed Figure 14 7 shows how a carbon dioxide laser is used (with a gas jet) in cutting wood Laser cutting of stainless steel, nickel alloys, and other metals finds widespread application in the aircraft and automobile industries

It has been recently tested and shown that aluminum sheet metal similar to that used in the aerospace industry can be efficiently cut with high-powered laser beam In fact, it is believed that it could be as much as 60% to 70% less expensive than the conventional techniques (*Optical Spectra*, September 1977, p 21)

Laser cutting has also been used in the textile industry for cutting cloth It is even claimed that this is the greatest advance in apparel manufacturing since the sewing machine

14 2 4 *Other Applications*

Lasers also find applications in vaporizing materials for subsequent deposition on a substrate Some unique advantages offered by the laser in such a scheme include the fact that no contamination occurs, some preselected areas of the source material may be evaporated, or the evaporant may be located very close to the substrate

Brittle materials like rock, marble, etc can be fractured using laser beams Such a technique finds application in rock crushing and boring tunnels

Lasers are also being used in the removal of microscopic quantities of material from balance wheels while in motion They have also been used

Fig 14 8 The photograph shows an oyster opened with a CO_2 laser, which neatly detach adductor muscle from the shell and leaves the raw oyster alive in the half shell (Photograph courtesy Professor Gurbax Singh of the University of Maryland)

in trimming resistors to accuracies of 0 1% Such micromachining processes find widespread use in semiconductor circuit processing The advantages offered by a system employing lasers for such purposes include the small size of the focused image with a precise control of energy, the absence of any contamination, accuracy of positioning, and ease of automation

An interesting application of the laser is in the opening of oysters A laser beam is focused on that point on the shell where the muscle is attached This results in detachment of the muscle, the opening of the shell, and leaving the raw oyster alive in the half shell (see Fig 14 8)

14 3 Laser Tracking

By tracking we imply either the determination of the trajectory of a moving object like an aircraft or a rocket, or determining the daily positions of a heavenly object (like the Moon) or an artificial satellite, a nice review on laser tracking systems has been given by Lehr (1974) The basic principle of laser tracking is essentially the same as that used in microwave radar systems In this technique, one usually measures the time taken to travel to and fro for a sharp laser pulse sent by the observer to be reflected by the object and received back by the observer (see Fig 14 9), suitably modulated cw (continuous wave) lasers can also be used for tracking

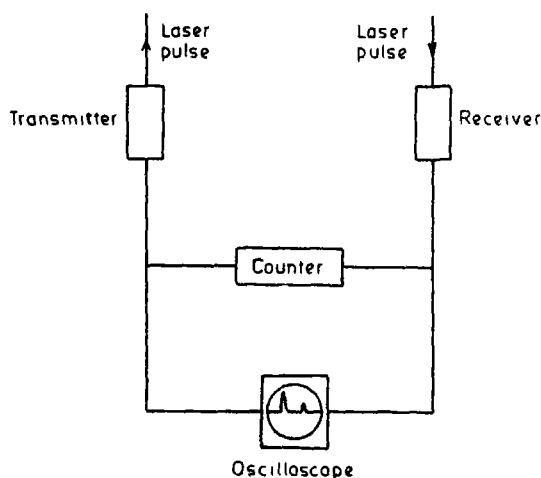



Fig 14 9 Block diagram for a tracking system using lasers

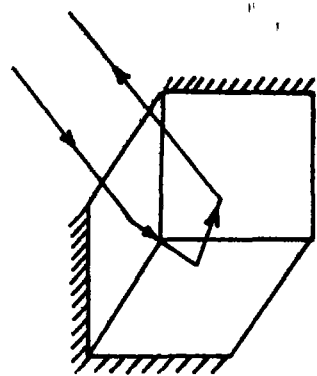



Fig 14 10 Cube corner as a retroreflector

One of the main advantages of a laser tracking system over a microwave radar system is the fact that a laser tracking system not only has a smaller size but its cost is also usually much less Further, in many cases one can use a retroreflector on the object, in a retroreflector, the incident and reflected rays are parallel and travel in opposite directions A cube corner is often used to act as a retroreflector (see Fig 14 10) For example, on the surface of the moon, or on a satellite, one can have a retroreflector to reflect back the incident radiation For a laser tracking system the size of the retroreflector is much smaller than the corresponding microwave reflector owing to the smaller wavelength of the optical beam and hence the reflector can be more conveniently mounted in the system involving lasers

In a microwave radar system, one has to incorporate corrections because of the presence of the ionosphere and also because of the presence of water vapor in the troposphere These corrections are much easier to incorporate in the case of an optical beam As compared to a microwave radar system, the laser radar offers much higher spatial resolution

On the other hand, there are some disadvantages in using a laser tracking system For example, when fog and snow are present in the atmosphere, it is extremely difficult to work at optical frequencies Further, during daytime there is a large background noise The losses in the transmitter and receiver are also considerably larger in laser systems

In Table 14 1 we have tabulated some of the typical lasers that have been used in tracking systems Typical ranges and velocities of various objects measured by a laser tracking system are tabulated in Table 14 2. One can see that the distance that can be covered range from 5000 m to hundreds of megameters

Table 14 1 Characteristics of Pulsed Lasers Used in Tracking Systems[a]

Type	Wave-length (μm)	Efficiency (%)	Energy (J)	Pulse duration (nsec)	Pulse repetition rate	Spectral width (nm)
Nd–YAG	1 06	0 1	0 02	10–25	100 sec^{-1}	0 5
GaAs	0 9	4	10^{-4}	100	100 sec^{-1}	2
Ruby	0 694	0 013	7	3	20 min^{-1}	0 04
Nd–glass	0 530	0 04	20	20	12 hr^{-1}	0 9

[a] Adapted from Lehr (1974)

The transmitter which is pointed towards the object may simply consist of a beam expander as shown in Fig 14 3 For tracking a moving object, both the laser and the telescope may be moved One could alternatively fix the laser and bend the laser beam by means of mirrors There are other ways of directing the laser beam towards the object, for further details, the reader is referred to the review article by Lehr (1974) and the references therein The receiver which is also pointed towards the object may consist of a reflector or a combination of mirrors and lenses The detector may simply be a photomultiplier

Figure 14 9 gives a block diagram of a laser radar system for tracking of a satellite A portion of the pulse that is sent is collected and is made to start an electronic counter The counter stops counting as soon as the reflected pulse is received back The counter may be directly calibrated in units of distance

Recently, the National Aeronautics and Space Administration, U.S A launched an aluminum sphere called the Laser Geodynamic Satellite (LAGEOS) into orbit at an altitude of 5800 km for studying the movements in the Earth's surface, which would be of great help in predicting earthquakes Figure 14 11 shows the satellite, which is 60 cm

Table 14 2 Typical Ranges and Velocities[a]

Object	Distance (m)	Angular velocity (arcsec sec^{-1})
Moon	$3\ 8 \times 10^{8}$	14 5
Near-Earth satellite	10^{6}	10^{3}
Aircraft (DC-10)	2×10^{4}	500
Rocket (at launch)	5×10^{3}	10^{5}

[a] Adapted from Lehr (1974)

Fig 14 11 The Laser Geodynamic Satellite (LAGEOS) put into orbit by the National Aeronautics and Space Administration, U S A for measuring minute movements of the Earth s crust which would be helpful in predicting earthquakes The satellite is 60 cm in diameter weighs 411 kg, and is studded with 426 retroreflectors, which return the incident laser pulses to their origin on the surface of the Earth Minute movements of the Earth s crust are detected by measuring the flight time of a light pulse to the satellite and back (Photograph courtesy United States Information Services New Delhi)

Fig 14 12 LAGEOS undergoing prelaunch testing in the laboratory (Photograph courtesy United States Information Service New Delhi)

in diameter, weighs 411 kg, and has 426 retroreflectors which return the laser pulses exactly back to the point of origin on the Earth Accurate measurements of the time of flight of laser pulses to the satellite and back should help scientists in measuring minute movements of the Earth's crust Figure 14 12 shows scientists performing the prelaunch testing of the satellite [see also *Optical Spectra* **13**(9), 21 (1979)]

14.4. Lidar

Laser systems have also been used for monitoring the environment Such systems are called lidars (acronym for light detection and ranging), and they essentially study the laser beam scattered from the atmosphere It may be mentioned that studies of the atmosphere using an optical beam had been carried out even before the advent of the laser, for example, using a searchlight, Hulbert in 1937 studied atmospheric turbidity to a height of 28 km The arrival of the laser on the scene revolutionized the atmospheric study using coherent light beams

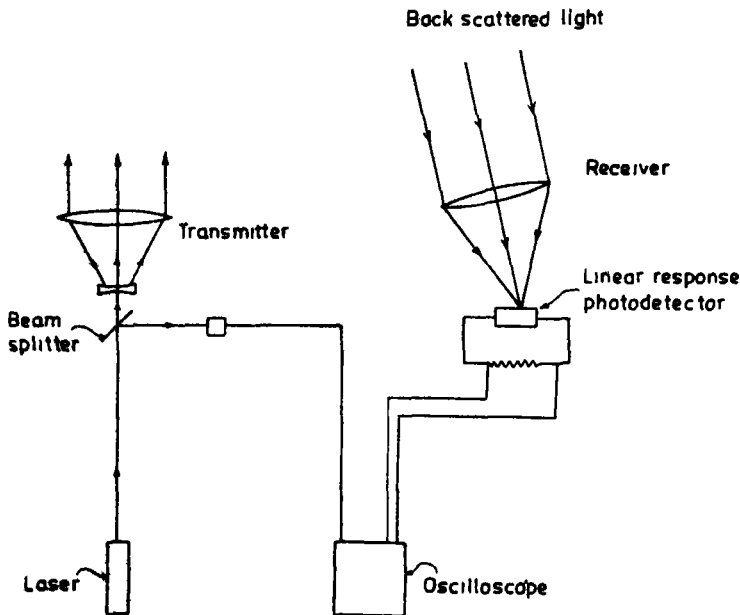



Fig 14 13 Block diagram of a pulsed lidar system A small fraction of the laser pulse is sent to the oscilloscope to record the zero time The backscattered light produces a current in the photodiode which is recorded on the oscilloscope

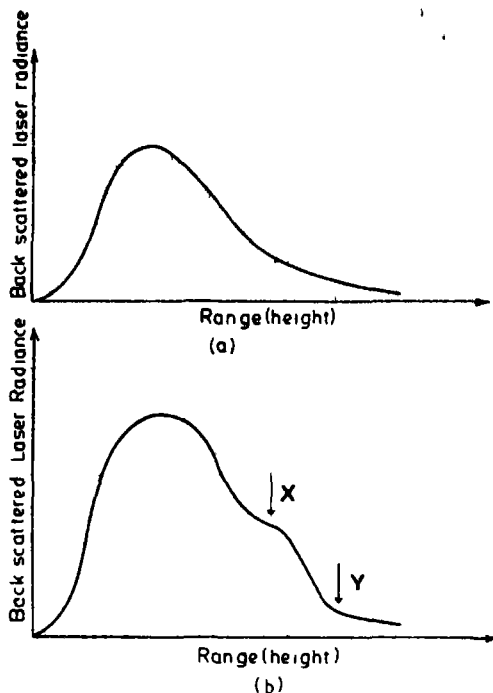



Fig 14 14 (a) Typical time dependence of the backscattered laser radiance as a function of height for a pure molecular atmosphere (b) Time dependence of the backscattered laser radiance as a function of height for an atmosphere containing aerosols Notice the kinks in the curve at points X and Y corresponding to the heights h_1 and h_2 (Adapted from Hall, 1974)

Pulses of laser light are sent and the radiation that is scattered by various particles present in the atmosphere is picked up by the receiver The background sunlight is removed by using filters This scattered light gives information regarding the particles present in the atmosphere with a sensitivity that is much more than that obtainable from microwave radars

In Fig 14 13 we have given a block diagram of a pulsed lidar system to study the nature of aerosols present in the atmosphere One usually measures the time dependence of the intensity of the backscattered laser light using a photodetector The time variation can be easily converted into the height from which the laser beam has been backscattered A typical time dependence of the backscattered laser radiance is plotted in Fig 14 14a, which corresponds to an atmosphere which has no aerosols, i e , the backscattering is by pure molecular gases such as N_2, O_2, Ar, etc † On the other hand, if the atmosphere contained aerosols, then the time dependence of the backscattered laser radiance would roughly be of

† It may be mentioned that the scattering from molecules of N_2, O_2 etc (whose dimensions are much smaller than the optical wavelength) is of the Rayleigh type, whereas the scattering from aerosols (whose dimensions are comparable to the laser wavelength) requires a more thorough analysis this is known as Mie scattering

the form shown in Fig 14 14b Notice the kinks that appear in the curve at the points marked X and Y, these are due to the fact that between the heights h_1 and h_2 there are aerosols which are responsible for a greater intensity (compared to that for a clear atmosphere) of the backscattered laser light Thus a curve like that shown in Fig 14 14b implies a haze which exists between the heights h_1 and h_2 It may be seen that corresponding to the height h_2 the intensity is roughly the same as that from a pure molecular atmosphere Thus, beyond the height h_2 one does not expect the presence of any aerosols With the lidar one can also study the concentrations and sizes of various particles present in the atmosphere, which are of extreme importance in pollution studies Small particles are difficult to detect with the microwave radar, the microwave radar can detect the presence of rain, hail, or snow in the atmosphere This difference arises essentially due to the larger amount of scattering that occurs at optical wavelengths In addition, a lidar can also be used to study the visibility of the atmosphere, the diffusion of particulate materials (or gases released at a point) in the atmosphere, and also to study the presence of clouds, fog, etc The study of turbulence and winds and the probing of the stratosphere have also been carried out by lidar systems For further details on the use of laser systems for monitoring the environment, the reader is referred to the review article by Hall (1974) and the references therein

14 5 Lasers in Medicine

Perhaps the most important use of lasers in the field of medicine is in eye surgery Hundreds of successful eye operations have already been performed using lasers The tremendous use of the laser in eye surgery is primarily due to the fact that the outer transparent regions of the eye allow light at suitable wavelengths to pass through for subsequent absorption by the tissues at the back of the eye

As is well known, the eye is roughly spherical and consists of an outer transparent wall called the cornea, which is followed by the iris (which can adjust its opening to control the amount of light entering the eye), and a lens Between the cornea and the lens is the aqueous humor The back part of the eye contains the light-sensitive element, namely, the retina Light falling on the eye is focused by the lens on the retina, and the photosensitive pigment-containing cells present in the retina convert the light energy into electrical signals, which are carried by the optic nerve to the brain, resulting in the process of seeing

As a result of some disease or heavy impact, the retinal layer may get

detached from the underlying tissue, creating a partial blindness in the affected area Earlier, a xenon-arc lamp was used for welding together the detached portion of the retina But the long exposure times of this source required administering anesthesia for safety Also it cannot be focused sharply

The unique advantages of using a laser beam for welding a detached retina are that since it can be focused to an extremely small spot, precise location of the weld can be made and also the welds are much smaller in size The spot size of a typical xenon-arc beam on the retina when focused by the eye lens is about 500–1000 μm in diameter, this is much larger than the typical diameter ($\sim 50\ \mu$m) obtainable using a laser beam The time involved in laser beam welding is so short that the eye does not need any clamping Pulses of light from a ruby laser lasting for about 300 μsec at levels below 1 J are used for retinal attachment

Lasers are also expected to be used extensively in the treatment of cancer In an experiment reported in the USSR, amelanotic melanoma was inculcated from human beings on nine animals These animals were irradiated with a ruby laser beam and it was reported that within one month the tumors completely disappeared The power associated with the ruby laser beam was about 100 MW with a total energy of about 200 J Successful skin cancer treatment with lasers have also been reported on human beings

It is impossible to list all the applications of lasers in the field of medicine Extensive use of lasers is anticipated in surgery, dentistry, and dermatology For further details and other applications of lasers in medicine, the reader is referred to the book by Goldman and Rockwell (1971)

14 6. Precision Length Measurement

The large coherence length and high output intensity coupled with a low divergence, enables the laser to find applications in precision length measurements using interferometric techniques The method essentially consists of dividing the beam from the laser by a beam splitter into two portions and then making them interfere after traversing two different paths (see Fig 14 15) One of the beams emerging from the beam splitter is reflected by a fixed reflector and the other usually by a retroreflector†

† As mentioned earlier a retroreflector reflects an incident beam in a direction exactly opposite to that of an incident beam (see Fig 14 10), and it is characterized by the property that minor misalignments of the moving surface do not cause any significant errors

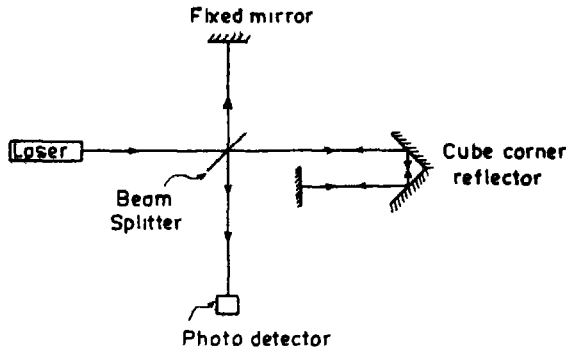



Fig. 14.15. Laser interferometer arrangement for precision length measurements

mounted on the surface whose position is to be monitored. The two reflected beams interfere to produce either constructive or destructive interference. Thus, as the reflecting surface is moved, one would obtain alternatively constructive and destructive interference, which can be detected with the help of a photodetector. Since the change from a constructive to a destructive interference corresponds to a change of a distance of half a wavelength, one can measure the distance traversed by the surface on which the reflector is mounted by counting the number of fringes which have crossed the photodetector. Accuracies up to 0.1 μm can be obtained by using such a technique.

This technique is being used for accurate positioning of aircraft components on a machine tool, for calibration and testing of machine tools, for comparison with standards, and many other precision measurements; for further details, see, e.g., Harry (1974), Chapter 5 and the references therein. The conventional cadmium light source can be used only over path differences of about 20 cm. With the laser one can make very accurate measurements over very long distances because of the large coherence length. The most common type of laser used in such applications is the helium–neon laser (see Section 9.4), and since the distance measurement is being made in terms of wavelength, in these measurements, a high wavelength stability of the laser output must be maintained.

14.7. Velocity Measurement

It is well known that when a light beam gets scattered by a moving object, the frequency of the scattered wave is different from that of the incident wave; the shift in the frequency depends on the velocity of the

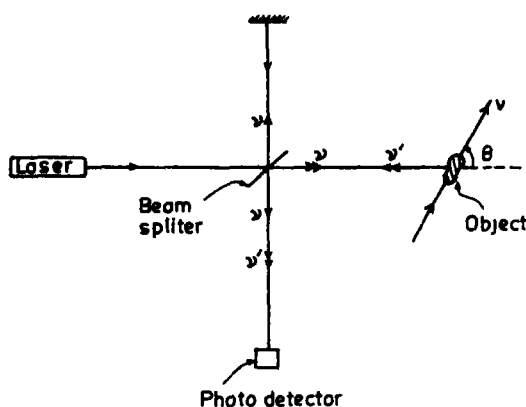



Fig 14 16 Schematic of an arrangement for measuring the velocity of a moving object using Doppler shift

object Indeed, if ν represents the light frequency and v represents the velocity of the moving object which is moving at an angle θ with respect to the incident light beam (see Fig 14 16), then the change in frequency $\Delta\nu$ between the incident and the reflected beams is given by

$$\frac{\Delta\nu}{\nu} = \frac{2v}{c}\cos\theta \qquad (14\ 7\text{-}1)$$

where c represents the velocity of light in free space Thus the change in frequency $\Delta\nu$ is directly proportional to the velocity v of the moving object, this is known as the Doppler shift Thus, by measuring the change in frequency suffered by a beam when scattered by a moving object, one can determine the velocity of the object This method has been successfully used for velocity determination of many types of materials from about 10 mm/min to about 150 m/min (Harry, 1974) Further, using the above principle, portable velocity-measuring meters have been fabricated which measure speeds in the range of 10–80 miles/hr, these have been used by traffic police Laser Doppler velocimeters have also been used for measuring fluid flow rates

The basic arrangement for velocity measurements is the following The beam from a cw laser (usually a helium–neon laser—see Section 9 4) is split by a beam splitter, one of the components is reflected back from a fixed mirror and the other component undergoes scattering from the moving object The two beams are then combined and made to interfere as shown in Fig 14 16, and because of the difference in frequency between the two beams, beating occurs The beat frequency is a direct measure of the velocity of motion of the object

PART III

The Nobel Lectures

Production of Coherent Radiation by Atoms and Molecules

Charles H. Townes

General and Historical Comments

From the time when man first saw the sunlight until very recently, the light which he has used has come dominantly from spontaneous emission, like the random emission of incandescent sources So have most other types of electromagnetic radiation—infrared, ultraviolet, or gamma rays The maximum radiation intensities, or specifically the power radiated per unit area per unit solid angle per unit frequency bandwidth, have been controlled by Planck's black-body law for radiation from hot objects This sets an upper limit on radiation intensity—a limit which increases with increasing temperature, but we have available temperatures of only a few tens of thousands, or possibly a few millions of degrees

Radio waves have been different And perhaps without our realizing it, even much of our thinking about radio waves has been different, in spite of Maxwell's demonstration before their discovery that the equations governing radio waves are identical with those for light The black-body law made radio waves so weak that emission from hot objects could not, for a long time, have been even detected Hence their discovery by Hertz and the great use of radio waves depended on the availability of quite different types of sources—oscillators and amplifiers, for which the idea of temperature and black-body radiation even seems rather out of place For example, if we express the radiation intensity of a modern electronic oscillator in terms of temperature, it would typically be in the range 10^{10}–10^{30} degrees Kelvin

These two regimes, radio electronics and optics, have now come

much closer together in the field known as quantum electronics, and have lent each other interesting insights and powerful techniques

The development of radar stimulated many important applications of electronics to scientific problems, and what occupied me in particular during the late nineteen-forties was microwave spectroscopy, the study of interactions between microwaves and molecules From this research, considerable information could be obtained about molecular, atomic, and nuclear structure For its success, coherent microwave oscillators were crucial in allowing a powerful high-resolution technique Consequently it was important for spectroscopy, as well as for some other purposes, to extend their range of operation to wavelengths shorter than the known limit of electronic oscillators, which was near one millimeter Harmonic generation and some special techniques allowed interesting, though rather slow, progress The basic problem with electronic amplifiers or oscillators seemed to be that inevitably some part of the device which required careful and controlled construction had to be about as small as the wavelength generated This set a limit to construction of operable devices It was this experimental difficulty which seemed inevitably to separate the techniques which were applicable in the radio region from those applicable to the shorter waves of infrared or optical radiation

Why not use the atomic and molecular oscillators already built for us by nature? This had been one recurring theme which was repeatedly rejected Thermodynamic arguments tell us, in addition to the black-body law of radiation, that the interaction between electromagnetic waves and matter at any temperature† cannot produce amplification For, radiation at the temperature of matter cannot be made more intense by interaction of the two without violating the second law But already by 1917, Einstein had followed thermodynamic arguments further to examine in some detail the nature of interactions between electromagnetic waves and a quantum-mechanical system And a review of his conclusions almost immediately suggests a way in which atoms or molecules can in fact amplify

The rate of change of electromagnetic energy confined in a region where it interacts with a group of molecules must, from Einstein's work, have the form

$$\frac{dI}{dt} = AN_b - BIN_a + B'IN_b \qquad (1)$$

where N_b and N_a are the numbers of molecules in the upper and lower of

† Strictly speaking, at any *positive* temperature Negative absolute temperatures can be defined as will be noted below

two quantum states, which we assume for simplicity to be nondegenerate (i.e single) A and B are constants, and thus the first and second terms represent spontaneous emission and absorption, respectively The third term represents emission from the upper state produced by the presence of a radiation intensity I, and is hence called stimulated emission

At equilibrium, when $dI/dt = 0$, $I = AN_b/(BN_a - B'N_b)$ Rather simple further thermodynamic reasoning shows that $B' = B$ and gives the ratio A/B While Boltzmann's law $N_b = N_a e^{-W/kT}$ requires $N_b < N_a$ at any temperature T, it is immediately clear from (1) and that if $N_b > N_a$, dI/dt will always be positive and thus the radiation amplified This condition is of course one of nonequilibrium for the group of molecules, which hence successfully obviates the limits set by black-body radiation The condition $N_b > N_a$ is also sometimes described as population inversion, or as a negative temperature [2] since in Boltzmann's law it may be obtained by assuming a negative absolute temperature

Thermodynamic equilibrium between two states of a group of atoms requires not only a Boltzmann relation $N_b = N_a e^{-W/kT}$, but also a randomness of phases of the wave functions for the atoms In classical terms, this means that if the atomic electrons are oscillating in each atom, there must not be a correlation in their phases if the entire group can be described as in temperature equilibrium Einstein's relation (1) in fact assumed that the phases are random And if they are not, we have another condition which will allow the atoms to amplify electromagnetic waves, even when N_b [illegible] This represents a second type of loophole in the limits set by the black body law and thermodynamic equilibrium, and one which can also be used alone or in conjunction with the first in order to produce amplification

Thermodynamic arguments can be pushed further to show that stimulated emission (or absorption) is coherent with the stimulating radiation That is the energy delivered by the molecular systems has the same field distribution and frequency as the stimluating radiation and hence a constant (possibly zero) phase difference This can also be shown somewhat more explicitly by a quantum-mechanical calculation of the transition process

Stimulated emission received little attention from experimentalists during the 1920's and 1930's when atomic and molecular spectroscopy were of central interest to many physicists

Later, in the 1940's experiments to demonstrate stimulated emission were at least discussed informally and were on the minds of several radio spectroscopists, including myself But they seemed only rather difficult demonstrations and not quite worthwhile In the beautiful 1950 paper of Lamb and Retherford on the fine structure of hydrogen [3], there is a

specific, brief note about "negative absorption" with reversal of popula tion And a year later, Purcell and Pound [4] published their strikin demonstration of population inversion and stimulated emission As matter of fact, population inversion and its effects on radiation ha already shown up in a somewhat less accented form in the resonanc experiments of Bloch [5] and others But all these effects were so smal that any amplification was swamped by losses due to other competin processes, and their use for amplification seems not to have been seri ously considered until the work of Basov and Prochorov [6], Weber [7] and of Gordon, Zeiger, and Townes [8, 9] in the early nineteen fifties

My own particular interest came about from the realization tha probably only through the use of molecular or atomic resonances coul coherent oscillators for very short waves be made, and the sudde discovery in 1951 of a particular scheme [8] which seemed to really offe the possibility of substantial generation of short waves by molecula amplification

Basic Maser Principles

The crucial requirement for generation, which was also recognized b Basov and Prochorov was to produce positive feedback by some reso nant circuit and to ensure that the gain in energy afforded the wave b stimulated molecular transitions was greater than the circuit losses Con sider a resonant microwave cavity with conducting walls a volume V, an a quality factor Q The latter is defined by the fact that power los because of resistance in the walls is $\overline{E^2}V\nu/4Q$, where $\overline{E^2}$ is the electri field strength in the mode averaged over the volume and ν is th frequency If a molecule in an excited state is placed in a particular fiel of strength E, the rate of transfer of energy to the field i $[E\mu/\hbar]^2(h\nu/3\,\Delta\nu)$ when the field's frequency coincides with the resonanc frequency ν between the two molecular states Here μ is a dipole matri element for the molecular transition and $\Delta\nu$ is the width of the molecula resonance at half maximum (assuming a Lorentz line shape) Hence fo N_b molecules in the upper state and N_a in the lower state, the powe given the field in the cavity is $(N_b - N_a)(E\mu/\hbar)^2(h\nu/3\,\Delta\nu)$ If th molecules are distributed uniformly throughout the cavity E^2 must b averaged over the volume For the net power gain to be positive, then

$$(N_b - N_a)\left(\overline{\frac{E\mu}{\hbar}}\right)^2 \frac{h\nu}{3\,\Delta\nu} \geqslant \frac{\overline{E^2}V\nu}{4Q}$$

This gives the threshold condition for buildup of oscillations in the cavity

$$(N_b - N_a) \geqslant \frac{3hV\,\Delta\nu}{16\pi^2 Q\mu^2} \tag{2}$$

There are by now an enormous variety of ways in which the threshold condition can be met, and some of them are strikingly simple But the system which first seemed to give an immediate hope of such an oscillator involved a beam of ammonia molecules entering a resonant cavity as shown in Fig 1 The transition used was the well-known inversion transition of ammonia at 23,870 Mc/sec A "focusser," involving inhomogeneous electric fields, tends to remove molecules in the ground state from the beam and to focus molecules in the excited state along the axis of the beam and into the cavity, thus ensuring that $N_b \gg N_a$ J P Gordon played a crucial role in making operable the first such system in 1954, after two and one-half years of experimental work [9, 10], and H J Zeiger was a valuable colleague in the first year of work and early designs We called this general type of system the maser, an acronym for *m*icrowave *a*mplification by *s*timulated *e*mission of *r*adiation The idea has been successfully extended to such a variety of devices

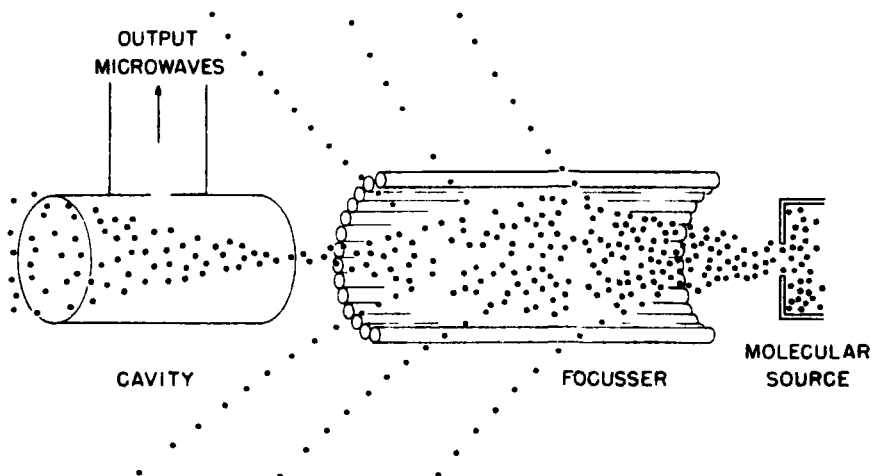



Fig 1 The ammonia (beam-type) maser Molecules diffuse from the source into a focusser where the excited molecules (open circles) are focussed into a cavity and molecules in the ground state (solid circles) are rejected A sufficient number of excited molecules will initiate an oscillating electromagnetic field in the cavity, which is emitted as the output microwaves Because of energy given to the field, some molecules return to the ground state towards the end of their transit through the cavity

and frequencies, that it is probably well to generalize the name—perha to mean *m*olecular *a*mplification by *s*timulated *e*mission of *r*adiation B in the radio-frequency range, it is sometimes called the *r*aser, and f light the term *l*aser is convenient and commonly used Maser amplific tion is the key process in the new field known as quantum electronics- that is, electronics in which phenomena of a specifically quantur mechanical nature play a prominent role

It is well known that an amplifier can usually be made into oscillator, or vice versa, with relatively minor modifications But it w only after experimental work on the maser was started that we realiz this type of amplifier is exceedingly noise-free The general reason for lc noise can be stated simply The molecules themselves are uncharged that their motions, in contrast to motions of electrons through vacuu tube amplifiers, produce no unwanted electromagnetic signals Hence signal introduced into the resonant cavity competes only with whatev thermal noise is in the cavity as the result of thermal radiation from tl cavity walls, and with spontaneous emission from the excited molecule Spontaneous emission can be regarded for this purpose as that stimulat by a fluctuating field of energy $h\nu$ Since $kT \approx 200\, h\nu$ for microwaves a cavity at room temperature, the thermal radiation kT in the cavity much more important than spontaneous emission It is then only tl thermal radiation present which sets the limit to background noise, sin it is amplified precisely as is the signal

The above discussion also shows that, if the cavity is at 0°K and extraneous noise enters the cavity with the input signal, the limiting noi fluctuation is determined by the spontaneous emission, which is equiv lent to only one quantum of energy in the cavity It can be shown, in fa that masers can yield the most perfect amplification allowed by tl uncertainty principle

The motion of an electromagnetic wave is analogous to that of mechanical harmonic oscillator, the electric and magnetic fields corr sponding to position and momentum of the oscillator Hence the quantur mechanical uncertainty principle produces an uncertainty in the simu taneous determination of the electric and magnetic fields in a wave, equivalently in determination of the total energy and phase of the wav Thus one can show that, to the extent that phase of an electromagne wave can be defined by a quantum-mechanical operator, there is uncertainty relation† [11]

$$\Delta n\, \Delta\Phi \geq 1/2 \qquad ($$

Here Δn is the uncertainty in the number of photons in the wave, and Δ is the uncertainty in phase measured in radians

† See also Section 8 6— Authors

Any amplifier which gives some representation of the phase and energy of an input wave in its output must, then, necessarily involve uncertainties or fluctuations in intensity Consider, for example, an ideal maser amplifier composed of a large number of molecules in the upper state interacting with an initial electromagnetic wave, which is considered the signal After some period of time, the electromagnetic wave will have grown to such magnitude that it contains a very large number of quanta and hence its phase and energy can be measured by classical means By using the expected or average gain and phase relation between the final electromagnetic wave and the initial signal, the maser amplifier thus allows a measurement of the initial wave

A calculation by well-established quantum-mechanical techniques of the relation between input and output waves shows that this measurement of the input wave leaves an uncertainty just equal to the minimum required by the uncertainty principle [11] Furthermore the product $\Delta E \Delta H$ of uncertainties in the electric and magnetic fields has the minimum value allowed while at the same time $(\Delta E)^2 + (\Delta H)^2$ is minimized The uncertainty in number n of quanta in the initial wave is $\Delta n = (n + 1)^{1/2}$, and in phase it is $\Delta\Phi = 1/2n^{1/2}$, so that $\Delta n\, \Delta\varphi = (1/2)[(n + 1)/n]^{1/2}$ The phase has real meaning, however, only when there are as many as several quanta, in which case $\Delta n\, \Delta\Phi \rightarrow 1/2$, the minimum allowed by (3) The background noise which is present with no input signal ($n = 0$), is seen to be equivalent to a single quantum ($\Delta n = 1$) of input signal

A somewhat less ideal maser might be made of N_b and N_a molecules in the upper and lower states, respectively, all interacting with the input signal In this case fluctuations are increased by the ratio $N_b/(N_b - N_a)$ If the amplifier has a continuous input signal, a continuous amplified output, and a bandwidth for amplification $\Delta\nu$, the noise power output can be shown to be equivalent to that produced by an input signal [12]

$$N = \frac{h\nu\, \Delta\nu}{1 - N_a/N_b}$$

The noise power N is customarily described in terms of the noise temperature T_n of the amplifier, defined by $N = KT_n\, \Delta\nu$ Thus the minimum noise temperature allowed by quantum mechanics is that for a maser with $N_a/N_b \ll 1$, which is

$$T_n = \frac{h\nu}{k} \tag{4}$$

This is equivalent to the minimum energy uncertainty indicated above of one quantum ($\Delta n = 1$) In the microwave region, T_n given by (4) is approximately 1°, whereas the best other microwave amplifiers when

maser amplifiers were first being developed had noise fluctuations about 1,000 times greater

It is interesting to compare an ideal maser as a detector with a perfect photodetector, such as a γ-ray counter The γ-ray counter can detect a single photon with almost no false signals, whereas a maser must always have a possible false signal of about one photon But the photodetector gives no information about the phase of the signal, it only counts quanta, which is why the uncertainty principle allows $\Delta n \rightarrow 0$ Unfortunately, there are no perfect photodetectors in the microwave or radio regions, so that the maser is our best available detector for these waves

The same freedom from noise which makes the maser a good amplifier helps make it a strikingly good source of monochromatic radiation, since when the threshold condition is fulfilled and the maser oscillates, the low noise implies a minimum of random frequency fluctuations

Consider now a maser oscillator consisting of a group of excited molecules in a resonant cavity Let the molecular transition frequency be ν_m, its half width at half-maximum intensity $\Delta\nu_m$, and the resonant cavity frequency be ν_c with a half width $\Delta\nu_c$ Assuming that ν_m and ν_c differ by much less than $\Delta\nu_m + \Delta\nu_c$, the radiation produced by the oscillation can be shown to occur at a frequency [13]

$$\nu = \frac{\nu_m Q_m + \nu_c Q_c}{Q_m + Q_c} \tag{5}$$

where the quality factors Q_m and Q_c are $\nu_m/\Delta\nu_m$ and $\nu_c/\Delta\nu_c$ respectively Thus if the molecular resonance is much sharper than that of the cavity, as in the ammonia beam maser ($Q_m \gg Q_c$), the frequency of oscillation is [10]

$$\nu = \nu_m + (\nu_c - \nu_m)\frac{Q_c}{Q_m} \tag{6}$$

If the cavity is tuned so that $\nu_c - \nu_m$ is small, then the frequency of oscillation coincides very closely with the natural molecular frequency ν_m, and one has an almost constant frequency oscillator based on a molecular motion, a so-called atomic clock

The frequency ν is not precisely defined or measurable because of noise fluctuations, which produce random phase fluctuations of the wave In fact, the maser is essentially like a positive feedback amplifier which amplifies whatever noise source happens to be present and thereby produces a more-or-less steady oscillation If Q_m or Q_c is high, and the amplifier gain is very large, then the bandwidth of the system becomes exceedingly small But it is never zero, nor is the frequency ever precisely

defined The average deviation in frequency from expression (5) which these phase fluctuations produce when averaged over a time t is [14]

$$\varepsilon = \Delta\nu\left(\frac{W_n}{Pt}\right)^{1/2} \tag{7}$$

where $\Delta\nu = \Delta\nu_c\, \Delta\nu_m/(\Delta\nu_c + \Delta\nu_m)$, P is the power generated by the oscillator, and W_n is the effective energy in the source of fluctuations Where $kT \gg h\nu$ in a cavity at temperature T and resonant frequency ν, the effective energy comes from thermal noise and $W_n = kT$ If the noise fluctuations come from spontaneous emission, as they do when $kT \ll h\nu$, then $W_n = h\nu$

It is also useful to state the spectral width of the radiation emitted from a maser oscillator, as well as the precision to which the frequency can be determined The half width of the spectral distribution is again determined by the same noise fluctuation and is given by [10, 15, 16]

$$\delta = \frac{2\pi W_n}{P}(\Delta\nu^2) \tag{8}$$

where $\Delta\nu$, W_n, and P are the same as in Equation (7) This width δ is typically so small in maser oscillators that they provide by far the most monochromatic sources of radiation available at their frequencies

Maser Clocks and Amplifiers

Although the ammonia beam-type maser was able to demonstrate the low-noise amplification which was predicted [17], its extremely narrow bandwidth makes it and other beam-type masers more useful as a very monochromatic source of electromagnetic waves than as an amplifier For the original maser, the power output P was about 10^{-9} watts, and the resonant width $\Delta\nu$ about 2 kc, as determined by the length of time required for the beam of molecules to pass through the cavity Since the frequency of oscillation ν_m is 23,874 Mc, the fractional spectral width according to (8) is $\delta/\nu \approx 10^{-14}$ In a time $t = 100$ sec, equation (7) shows that the frequency can be specified to a fractional precision $\varepsilon/\nu = 2 \times 10^{-14}$, and of course the precision increases for longer times proportionally to $1/t^{1/2}$

As a constant frequency oscillator or precise atomic clock, however, the ammonia maser has an additional problem which is not so fundamental, but which sets a limit on long-term stability This comes from long-term drifts, particularly of the cavity temperature, which vary ν_c These variations can be seen from (6) to "pull" ν Variations of this type

have limited the long-term stability [18] of ammonia masers to fractional variations of about 10^{-11}—this still represents a remarkably good clock

A beam-type maser using the hyperfine structure transition in the ground state of hydrogen, which is at 1,420 Mc, has recently been developed by Goldenberg, Kleppner, and Ramsey [19] In this case, the excited atoms bounce many times from glass walls in the cavity and thereby a resonance width as small as 1 c p s. is achieved Present designs of the hydrogen maser yield an oscillator with long-term fractional variations no larger than about 10^{-13} This system seems likely to produce our best available clock or time standard

Masers of reasonably wide utility as amplifiers came into view with the realization that certain solids containing paramagnetic impurities allowed attainment of the maser threshold condition [20]. Microwave resonances of paramagnetic atoms in solids, or in liquids, had been studied for some time, and many of their properties were already well known The widths of these resonances vary with materials and with impurity concentration from a small fraction of a megacycle to many hundreds of megacycles, and their frequencies depend on applied magnetic field strengths, so that they are easily tunable Thus they offer for maser amplifiers a choice of a considerable range of bandwidth, and a continuous range of frequencies

A paramagnetic atom of spin 1/2 has two energy levels which, when placed in a magnetic field, are separated by an amount usually about $\nu = 2.8H$ Mc Here H is the field in gauss, and from this it is clear that most of the microwave frequency range can be covered by magnetic fields of normal magnitudes The first paramagnetic masers suggested involved impurity atoms of this type in crystals of Si or Ge Relaxation between the two states was slow enough in these cases that a sufficient population inversion could be achieved [20] However, before very long a very much more convenient scheme for using paramagnetic resonances was proposed by Bloembergen [21], the so-called three-level solid-state maser This system allowed continuous inversion of population and hence continuous amplification which was very awkward to obtain in the previous two-level system

Paramagnetic atoms with an angular momentum due to electron spin S greater than 1/2 have $2S + 1$ levels which are degenerate when the atom is in free space But these levels may be split by "crystalline fields," or interaction with neighboring atoms if the atoms are imbedded in a solid, and frequently the splittings lie in the microwave range The energy levels of such a system, involving a spin of 3/2 and four levels, can be as indicated in Fig 2 when the system is in a magnetic field If a sufficiently large electromagnetic wave of frequency ν_{13} (the transition frequency

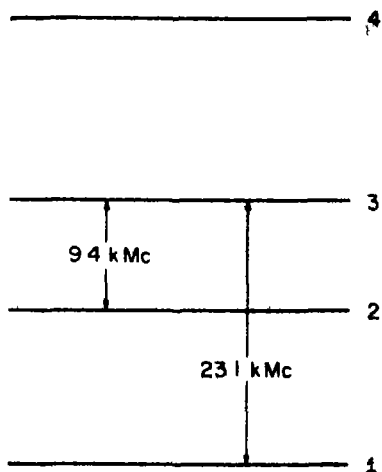



Fig 2 Energy levels of Cr^{+++} in ruby with a particular crystalline orientation in a magnetic field of 3,900 Oersteds For a three-level maser 23 1 kMc is the frequency of the pumping field and 9 4 kMc the frequency of amplication or oscillation

between levels 1 and 3) is applied, the population of these two levels can be equalized or "saturated " In this case, the ratio of the population of level 2 to that of 1 or 3 under steady conditions is

$$\frac{n_2}{n_1} = \frac{(1/T_{12})\exp(-h\nu_{12}/kT) + 1/T_{23}}{1/T_{12} + (1/T_{23})\exp(-h\nu_{23}/kT)}$$

Here T is the temperature of the crystal containing the impurities, and T_{12} and T_{23} are the times for relaxation between the states 1 and 2 or 2 and 3 respectively For $h\nu_{12} \gg kT$ and $h\nu_{23} \gg kT$, as occurs at very low temperatures or at ordinary temperatures if the levels are separated by optical frequencies,

$$\frac{n_2}{n_1} = \frac{T_{12}}{T_{23}}$$

When $h\nu_{12} \ll kT$ and $h\nu_{23} \ll kT$, which is more commonly the case for microwaves,

$$\frac{n_2}{n_1} = 1 + \frac{h}{kT}\frac{(\nu_{12}/T_{12}) - (\nu_{23}/T_{23})}{(1/T_{12}) + (1/T_{23})} \tag{9}$$

Thus if $\nu_{12}/T_{12} > \nu_{23}/T_{23}$, there is an inversion of population between levels 2 and 1, or if $\nu_{12}/T_{12} < \nu_{23}/T_{23}$ there is an inversion of population between levels 3 and 2 since the populations n_3 and n_1 have been equalized by the 'pumping" radiation Equation (9) is essentially the result obtained by Bloembergen [21] who also suggested several promising paramagnetic materials which might be used Basov and Prochorov

had already proposed a rather similar three-level "pumping" scheme for application to a molecular beam system [22]

The first successful paramagnetic maser of this general type was obtained by Scovil *et al* [23], using a rare earth ion in a water soluble crystal But before long, other more suitable crystals such as ruby [24] (chromium ions in Al_2O_3) became more or less standard and have provided amplifiers of remarkable sensitivity for radio astronomy, for satellite communication, and for communication to space probes [25] They have considerably improved the potentialities of radio astronomy, and already led to some new discoveries [26, 27] These systems generally require cooling with liquid helium, which is a technological difficulty that some day may be obviated But otherwise, they represent rather serviceable and convenient amplifiers

A maser amplifier of microwaves can rather easily be built which has a theoretical noise temperature as low as 1° or 2°K, and experimental measurements have confirmed this figure [28] However, such a low noise level is not easy to measure because almost any measurement involves attachment of input and output circuits which are at temperatures much higher than 1°K, and which radiate some additional noise into the amplifier The lowest over-all noise temperature so far reported for an entire receiving system [29] using a maser amplifier is about 10°K This represents about 100 times the sensitivity of microwave amplifiers which had been built before invention of the maser But masers have stimulated other amplifier work, and some parametric amplifiers, using more-or-less classical properties of materials rather than quantum electronics, now have sensitivites within a factor of about 5 of this figure

Optical and Infrared Masers, or Lasers

Until about 1957, the coherent generation of frequencies higher than those which could be obtained from electronic oscillators still had not been directly attacked, although several schemes using molecular-beam masers for the far infrared were examined from time to time This lack of attention to what had been an original goal of the maser came about partly because the preliminary stages, including microwave oscillators, low-noise amplifiers, and their use in various scientific experiments, had proved so interesting that they distracted attention from the high-frequency possibilities

But joint work with A L Schawlow [30] beginning at about this time helped open the way for fairly rapid and interesting development of maser oscillators in the far infrared, optical, and ultraviolet regions—as

much as 1,000 times higher in frequency than any coherent sources of radiation previously available It is masers in these regions of the spectrum, frequently called lasers for *l*ight *a*mplification by *s*timulated *e*mission of *r*adiation, which have perhaps provided the most striking new scientific tools and results Important aspects of this work were clear demonstrations that there are practical systems which can meet the threshold condition of oscillation, and that particular resonator designs allow the oscillations to be confined to certain specific and desirable modes The resonator analyzed was composed simply of two parallel mirrors—the well-known Fabry–Perot interferometer, but of special dimensions

For light waves, the wavelength is so short that any macroscopic resonator constructed must have dimensions large compared to the wavelength In this case, the electromagnetic field may to some resonable approximation be considered to travel in straight lines and be reflected from the walls of the resonator The threshold condition may be written

$$\left(\frac{\mu E}{\hbar}\right)^2 \frac{h\nu(N_b - N_a)}{12\pi\,\Delta\nu} \geqslant \frac{E^2}{8\pi}\frac{V}{t} \tag{10}$$

where t is the decay time for the light in a cavity of reflecting walls and volume V If the light has a random path in the cavity, the decay time can be expressed generally in terms of the reflection coefficient r of the walls, the volume V, the wall area A, and the velocity of light c

$$t = \frac{6V}{(1 - r)Ac}$$

Hence (10) becomes [30]

$$N_b - N_a \geqslant \frac{\Delta\nu}{\nu}\frac{h(1 - r)Ac}{16\pi^2\mu^2 V} \tag{11}$$

It can be seen that this critical condition is almost independent of frequency if the fractional line width $\Delta\nu/\nu$ does not change with frequency (as, for example, in the Doppler effect) The reflection coefficient and dipole moment matrix element μ are not particularly dependent on frequency over the range in question Hence, if the critical condition can be met for one frequency, it can probably be met over the entire range from the far infrared to the ultraviolet

There is a problem with a resonator which is large compared to a wavelength in that there are many modes Hence unless the modes in which oscillations occur are successfully controlled, the electromagnetic field may build up simultaneously in many modes and at many frequencies The total number of modes in a cavity with frequencies which lie

within the line width $\Delta\nu$ of the atomic molecular resonance is

$$p = 8\pi \frac{V\nu^2 \Delta\nu}{c^3}$$

or about 10^9 for a cavity volume of 1 cm^3, a frequency in the optical region, and ordinary atomic line widths But fortunately, the possibility of oscillation can be eliminated for most of these modes

Two small parallel mirrors separated by a distance much larger than their diameter will allow a beam of light traveling along the axis joining them to travel back and forth many times For such a beam, the decay time t is $(L/c)(1 - r)$, where L is the mirror separation and r the reflectivity Hence the threshold condition is

$$N_b - N_a \geq \frac{3\,\Delta\nu}{8\pi^2\nu}\frac{hc(1 - r)}{\mu^2 L}$$

This assumes that diffraction losses are negligible A beam of light which is not traveling in a direction parallel to the axis will disappear from the volume between the mirrors much more rapidly Hence the threshold condition for off-axis beams will require appreciably more excited atoms than that for axial beams, and the condition for oscillation can be met for the latter without a build-up of energy in off-axis light waves

Many features of the modes for the electromagnetic wave between two square plane parallel mirrors of dimension D and separation L can be approximately described as those in a rectangular box of these dimensions, although the boundary conditions on the enclosed sides of the "box" are of course somewhat different The resonant wavelengths of such a region for waves traveling back and forth in a nearly axial direction are [30]

$$\lambda = \frac{2L}{q}\left[1 - \frac{1}{2}\left(\frac{L}{D}\frac{r}{q}\right)^2 - \frac{1}{2}\left(\frac{L}{D}\frac{s}{q}\right)^2\right] \tag{12}$$

where q, r, and s are integers, and $r \ll q$, $s \ll q$ More precise examination of the modes requires detailed numerical calculation [31] For a precisely axial direction, $r = s = 0$, and the modes are separated by a frequency $c/2L$ If this frequency is somewhat greater than the atomic line width $\Delta\nu$, then only one axial mode can oscillate at a time The axial wave has an angular width due to diffraction of about λ/D, and if this is comparable with the angle D/L, then all off-axis modes (r or $s \neq 0$) are appreciably more lossy than are the axial ones, and their oscillations are suppressed

If one of the mirrors is partially transparent some of the light escapes from the axial mode in an approximately plane wave and with an angular divergence approximately λ/D determined by diffraction

A number of modified resonator designs have been popular and useful in optical masers, in particular ones based on the confocal Fabry–Perot interferometer However, the plane-parallel case seems to offer the simplest means of selecting an individual mode

Although a number of types of atomic systems and excitation seemed promising in 1958 as bases for optical masers, optical excitation of the alkali vapors lent itself to the most complete analysis and planning for an operable oscillator One such system has been shown to oscillate as expected [32], but the alkali vapors are no longer of great interest, because other systems which were at the time much less predictable have turned out to be considerably more useful

The first operating laser, a system involving optical excitation of the Cr ions in ruby and yielding red light, was demonstrated by Maiman in 1960 [33] He took what seemed at first a rather difficult route of inverting the population between the ground state and excited states of the Cr ion This technique requires that at least half of the very large number of atoms in the ground state must be excited in order to have the possibility of a population inversion In the case of two normally unpopulated atomic states, the total amount of excitation required is much less However Maiman succeeded handsomely in exciting more than half the Cr ions in a ruby of about 1/2,000 Cr concentration by applying a very intense pulse of light from a flash tube This type of system is illustrated schematically in Fig 3 Success immediately yielded a very high energy maser oscillation, because to get population inversion at all, a large amount of energy must be stored in the excited atomic states Surfaces of the ruby served as the reflecting mirrors Collins *et al* [34] quickly demonstrated that the ruby laser showed many of the characteristics predicted for such an oscillator

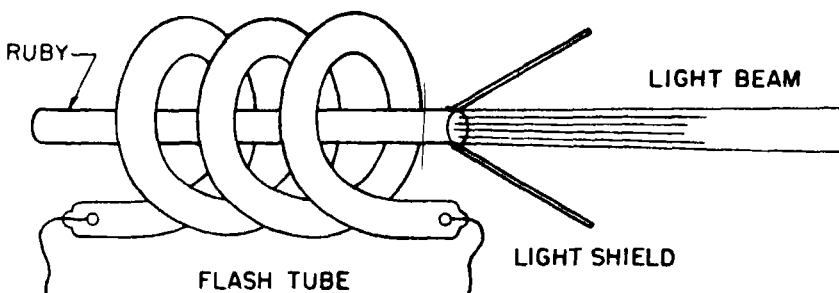



Fig 3 Schematic of a ruby (optically excited solid-state) laser When the gas flash tube is activated electromagnetic oscillations occur within the ruby rod, and some of these light waves are emitted in a beam through one partially reflecting end of the rod

The ruby laser is operated normally only in pulses, because of the high power required to reach threshold, and emits intense bursts of red light at power levels between about 1 kilowatt and 100 megawatts It has given rise to a whole family of lasers involving impurities in various crystals of glasses, and covering frequencies from the near infrared into the optical region

Not very long after the ruby laser was developed, Javan, Bennett, and Herriott [35] obtained maser oscillations from Ne atoms excited by collisions of the second kind with metastable He, in accordance with an idea previously put forward by Javan [36] This system, illustrated in Fig 4, requires only a gaseous discharge through a tube containing a mixture of He and Ne at low pressure, and two reflectors at the ends of the tube It oscillates at the relatively low power of about one milliwatt, but approaches ideal conditions much more closely than the ruby system, and affords a continuous source of infrared radiation of great purity and directivity

The technique of gaseous excitation by electrical discharge has also led to a large family of lasers, producing hundreds of different frequencies from many different gases which range from wavelengths as long as a few tenths of one millimeter down into the ultraviolet For some systems, a heavy discharge pulse in the gas is needed Others, particularly some of

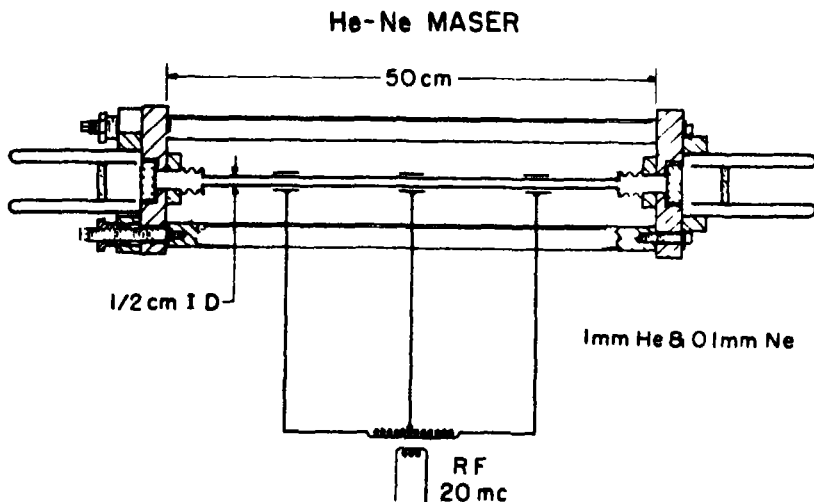



Fig 4 Schematic of a He–Ne (gas discharge) laser Electrical excitation can initiate a steady maser oscillation resulting in an emitted light beam from either end of the gas discharge, where there are reflecting mirrors

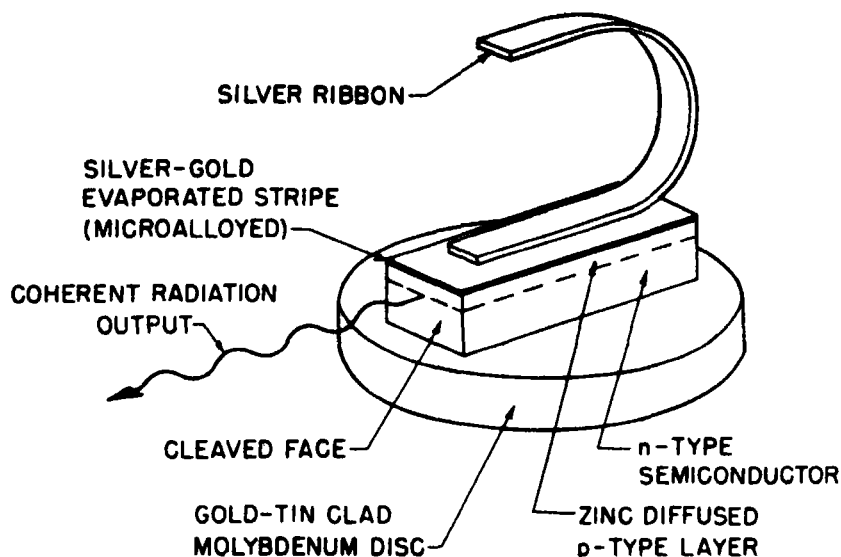



Fig 5 Schematic of a gallium arsenide (injection, or semiconductor) laser A small voltage applied between the silver ribbon and the molybdenum disc can produce maser oscillations with resulting emission of coherent infrared radiation

the infrared frequencies in rare gases, oscillate so readily that it seems probable that we have had lasers accidentally all along Very likely some neon or other rare gas electric signs have been producing maser oscillations at infrared wavelengths which have gone unnoticed because the infrared could not escape from the glass neon tubes Some of these oscillation frequencies represent atomic transitions which were previously undetected, for others the transition has not yet even been identified

Another class of lasers was initiated through the discovery [37] that a *p–n* junction of the semiconductor GaAs through which a current is passed can emit near infrared light from recombination processes with very high efficiency Hall *et al* [38] obtained the first maser oscillations with such a system, with light traveling parallel to the junction and reflected back and forth between the faces of the small GaAs crystal His results were paralleled or followed immediately, however, by similar work in two other laboratories [39, 40] This type of laser, illustrated in Fig 5, is of the general size and expense of a transistor It can be made to oscillate simply by passage of an electric current, and in some cases the radiation emitted represents more than 50% of the input electrical energy—an efficiency greater than that of other man-made light sources

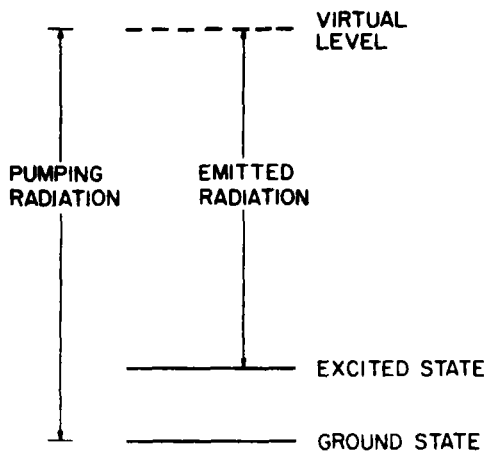



Fig 6 Representation of energy levels in a Raman maser This system resembles qualitatively a three-level maser with one of the levels being virtual or not characteristic of the molecule when no field is present

There quickly developed a large family of semiconductor lasers, some involving junctions and recently some using excitation by an external beam of electrons [41] They range in wavelength from about 10 microns in the infrared to the center of the visible region

Normal Raman scattering can be regarded as spontaneous emission from a virtual state, as indicated in Fig 6 Associated with any such spontaneous emission there must be, in accordance with Einstein's relations, a stimulated emission Javan showed [42] the principles involved in using this stimulated emission for a Raman maser What is required is simply a large enough number of molecular systems which are sufficiently strongly excited by radiation of frequency greater than some Raman-allowed transition

One might consider the population of the virtual level in a Raman maser (see Fig 6) to be greater than that of the first excited state, so that there is no population inversion On the other hand the initial state, which is the ground state, needs to be more populated that the first excited state One can quite properly consider the amplification process as a parametric one with the molecular frequency as idler, or as due to a mixture of ground and excited states in which there is phase coherence between the various molecules This is the second type of loophole through the black-body radiation law mentioned earlier The ammonia-beam maser itself illustrates the case of amplification without the necessity of population inversion As the ammonia molecules progress through the cavity and become predominantly in the ground state rather than the excited state, they continue to amplify because their oscillations are

correlated in phase with each other, and have the appropriate phase with respect to the electromagnetic wave

Raman masers were first demonstrated by Woodbury and Ng [43] as the result of excitation of various liquid molecules with a very intense beam from a pulsed laser They too have now many versions, giving frequencies which differ from the original driving maser beam by some small integer times a molecular-vibrational frequency Their action has been considerably extended by Terhune [44], and treated in a number of theoretical papers [42, 45]

Present Performance of Lasers

Where now do we stand in achieving the various theoretical expectations for performance of masers?

First, consider the general extension of the frequency range where we have coherent amplifiers and oscillators This has been increased by a factor somewhat more than 1,000, there are still additional spectral regions where such techniques need to be developed, but the pace has been quite rapid in the last few years Maser oscillations in the infrared, optical, or ultraviolet regions have now been obtained in many ways and appear easy, new excitation mechanisms and systems are continually turning up There are still two frequency regions, however, where such sources of radiation are rare or nonexistent One is in the submillimeter or far infrared The region has, in a sense, now been crossed and conquered by maser oscillators But techniques in this spectral region are still rudimentary and the frequency coverage with masers is spotty Presumably further work will allow interesting explorations in this region and a very fruitful high-resolution spectroscopy

Another region where coherent oscillators have not yet been developed is that of still shorter wavelengths stretching indefinitely beyond the near ultraviolet, where the first such oscillators are now available It can be shown that a rather severe and fundamental limitation exists as one proceeds to shorter wavelengths because of the continually increasing number of electromagnetic modes in a given volume and the faster and faster dissipation of energy into them by spontaneous emission

Consider a cavity resonator of fixed volume, fixed-wall reflectivity, and fixed-fractional frequency width $\Delta\nu/\nu$ Meeting the threshold condition (11) in such a resonator requires that there is power which increases as ν^4 radiated by spontaneous emission into all modes of the system [30] In the optical region this power dissipated amounts to only a few

milliwatts for typical conditions, whereas at 50 Å, in the soft X-ray region, it would be about 10^5 watts The threshold condition would then be very difficult to maintain But by the same token, if it is maintained the coherent X-ray beam produced would contain many kilowatts of power It seems reasonable to expect on this basis for masers to be developed to wavelengths somewhat below 1,000 Å, but that maser oscillations in the X-ray region will be very much more difficult

Secondly, let us examine the monochromaticity which has been achieved For the ammonia-beam maser, the phase variation of micro wave oscillations was shown experimentally to agree with the theoreti cal expression (7) within the experimental precision of about 50% This was done by beating two independent ammonia oscillators together and examining their relative phase variations [46] A similar procedure can be carried out for two optical oscillators by mixing their two light beams together in a photocell, and detecting the beat frequency However the technical difficulties in obtaining theoretical performance are rather more demanding than in the case of the ammonia maser Expression (8) for a typical He–Ne laser predicts a frequency spread of about 10^{-2} c p s or a fraction 3×10^{-17} of the oscillation frequency of 3×10^{14} c p s

Almost all masers so far oscillating in the optical or near infrared region require a sharper resonance, or higher Q, of the cavity than that of the atomic resonance Hence the frequency of oscillation is primarily determined, from (5), by the cavity resonance The frequency of oscillation thus depends on the separation L between mirrors, since from (12) $\nu = qc/2L$, where q is some integer If then, the radiated frequencies are to have a fractional bandwidth of about 3×10^{-17}, as would come from fundamental noise according to (8), the mirror separation must not vary by more than this fractional amount For a mirror separation of one meter, the motion allowed would be less than 3×10^{-15} cm—a demanding requirement!

If the mirror separation is held constant by cylindrical rods, L must still vary as a result of thermal excitation of the lowest frequency-stretching modes of the rods This gives an additional fluctuation which is usually larger than that from spontaneous emission It produces a fractional motion [47] $(2kT/YV)^{1/2}$, where T is the temperature, V the volume of the separators, and Y their Young's modulus

In order to examine the monochromaticity of lasers, two He–Ne systems were carefully shock-mounted in an acoustically insulated wine cellar of an unoccupied and isolated house so that acoustic vibrations would be minimized [47] Their pairs of mirrors were separated by heavy Invar rods about 60 cm long For this case, the limiting theoretical fluctuations set by thermal motions of the rods corresponded to fractional

frequency variations of 5×10^{-15}, or a frequency fluctuation of 2 c.p s Light from each laser was sent into a photodetector, and the beat frequency examined electronically Under good conditions free from acoustic disturbances or thermal transients in the Invar spacers, this experiment showed that variations in the laser frequencies over periods of a few seconds were less than 20 c p s , or about one part in 10^{13} This was ten times the limit of thermal fluctuations, but corresponded to detection of motions of the two mirrors as small as 5×10^{-12} cm, a dimension comparable with nuclear diameters Presumably, great care can obtain results still nearer to the theoretical

The narrowest atomic spectral lines have widths of the order of 10^8 c p s , so that the laser measured was more monochromatic than earlier light sources by about a factor of 10^6 Light of this type can interfere with itself after traveling a distance of about 10,000 km Hence it could in principle measure changes in such a large distance to a precision of one wavelength of light, assuming there were any optical path so constant Interference work has been done in several laboratories with laser light over distances of a few hundred meters, which does not require quite such special elimination of acoustic or other disturbances

A third property of laser light which is of interest is its directivity, or the spacial coherence across the beam As indicated above, certain modes of oscillation should represent approximately a plane wave of cross-section comparable with the mirror diameter D The He–Ne maser seems to easily allow adjustment so that such a mode of oscillation occurs, and its beam has been shown [35, 48] to have nearly the expected divergence λ/D due to diffraction

The spacial coherence or planarity of a laser beam implies that the entire beam can be focussed by a microscope to a region as small as about $\lambda/2$, or the resolving power of the microscope Similarly, it may be transmitted through a telescope in a beam whose angular width is simply determined by the angular resolution of the telescope and hence much less than the angular divergence λ/D as the beam emerges from a small laser The entire energy is originally created in the ideal laser in a single mode, it can be transmitted into other single modes by optical systems without violating the well-known brightness laws of optics

This brings us to a fourth important property, the intensity or brightness which can be achieved by maser techniques As indicated initially, once one has the possibility of coherent amplification, there is no firm limit to intensity because equilibrium thermodynamics and Planck's law no longer are controlling The only limit is set by the available energy input, heat dissipation, and size of the apparatus used

If only the one milliwatt of power emitted by a He–Ne laser is

focussed by a good lens, the power density becomes high because the cross-sectional area of the focussed spot would be only about $\lambda^2/4$ This gives a power density of 4×10^5 watts/cm^2 The effective temperature of such a beam, because of its monochromaticity, is also rather high—about 10^{19}°K for the 20 c p s bandwidth light

The pulsed systems, such as ruby lasers in particular, emit much greater power, although they do not quite approach the limits of coherence which the gaseous systems do Ruby lasers emit a few tenths to a few hundred joules of energy in pulses from about 10^{-3} sec to 10^{-8} sec in length The power can thus be as great as 10^9 watts or more Effective temperatures of the radiation are of the order 10^{23}°K The actual limit of power density will generally be set by the limit of light intensity optical materials can stand without breakage or ionization 10^9 watts focussed to a spot 10^{-2} mm in diameter produces an electric field strength in the optical wave of about 10^9 volts/cm, which is in the range of fields by which valence electrons are held in atoms Hence this power ionizes and disrupts all material The radiation pressure also becomes large, being about 10^{12} dynes/cm^2 or 10^6 atmospheres at such a focal point

Some Applications of Lasers

It is clear that light in more ideal and in more intense form, which maser techniques have produced, can be expected to find application in wide and numerous areas of technology and of science simply because we find our present techniques of producing and controlling light are already so widely applied Most of these applications are still ahead of us, and there is not time to treat here even those which are already beginning to develop I shall only mention that in technology, lasers have been put to work in such diverse areas as radar, surgery, welding, surveying, and microscopy A little more space will be devoted here to discussing three broad areas of science to which optical, infrared, and ultraviolet masers are expected to contribute

Masers seem to provide the most precise techniques for measurement of the two fundamental dimensions of time and length Over short periods of time, maser oscillators clearly give the most constant oscillations, for longer times the hydrogen maser also seems to provide the most precise clock yet available Light from optical masers allows new precision in the measurement of distance, and already seems capable of improving our standard of length This new precision suggests interesting experiments on certain fundamental properties of our space, as well as

the application of higher precision to a variety of physical effects So far, experiments have been done to improve the precision with which the Lorentz transformation can be experimentally verified [49, 50] It appears that improved precision in measurement of the speed of light can also be expected If we look some distance in the future, it seems clear that the techniques of quantum electronics will allow direct measurement of the frequency of light, rather than only its wavelength This can be accomplished by generation of harmonics of a radio frequency, amplification of the new frequency, and further generation of harmonics until the radio region is linked with optical frequencies This should eventually allow measurement of the velocity of light c to whatever precision we define time and length Or, it will allow the elimination of separate standards of time and of length because c times a standard time will define a standard length with more precision than we can now achieve

The power of spectroscopy should be considerably increased by use of masers In particular, these very monochromatic sources can very much improve spectroscopic resolution and thus allow more detailed examination of the structure of atoms, molecules, or solids This advance can be particularly striking in the infrared and far infrared where present resolution is far less than the widths of atomic or molecular lines Already some high-resolution spectroscopy has been done with lasers [51, 52], and still more interesting work of this general type can be expected before long

A third interesting field for which lasers are important has emerged as a field almost entirely because of the existence of lasers, and is the area where scientific research has so far been most active This is what is usually called nonlinear optics [53, 54], although it includes some phenomena which might not previously have been described this way We have been accustomed in the past to discussing the progress of light through a passive optical material of more-or-less fixed properties But in the intense laser beams now available, interactions between the light and the optical medium are sufficiently large that properties of the medium can no longer be regarded as fixed The medium distorts, its molecules vibrate, and polarization of electrons in its atoms no longer responds linearly to the applied field One must now also consider the dynamics of both the light and the optical medium, and interactions between their two motions Some of the new phenomena observed are multiple-quanta absorption, which makes absorption depend on intensity [55, 56], harmonic generation in optical materials and mixing of light frequencies [57, 58, 59, 60], excitation of coherent molecular vibrations and stimulated Raman effects [42, 43, 44, 45], and stimulated Brillouin scattering [61, 62] Only the last two of these will be discussed, partly because they bear on still another kind of maser, one which generates phonons

The Phonon Maser

Acoustic waves follow the same general form of equations as do light waves and manifest many of the same phenomena An acoustic wave can produce an atomic or molecular excitation, or receive energy from it by either spontaneous or by stimulated emission Hence, one may expect maser action for acoustic waves if a system can be found in which molecules are sufficiently coupled to an acoustic field and appropriate excitation can be obtained to meet the threshold condition The first such systems suggested involved inversion of the spin states of impurities in a crystal in ways similar to those used for solid-state electromagnetic masers [63] A system of this type has been shown to operate as expected [64] However, a more generally applicable technique seems to be Brillouin scattering, and its close associate Raman scattering, which utilizes phase correlation rather than population inversion to produce amplification This process can also be viewed as parametric amplification

Light may be scattered by the train of crests and troughs in an acoustic wave much as by a grating Since the wave is moving, the scattering involves a Doppler shift The net result, first analyzed by Brillouin [65], is that the scattered light is shifted in frequency from the frequency ν_0 of the original beam by an amount

$$\nu = 2\nu_0 v/c \sin \theta/2 \qquad (13)$$

where v and c are the phase velocities of sound and of light, respectively, in the medium, and θ is the scattering angle The energy lost $h\nu$ is given to the scattering acoustic wave of frequency ν If the light is of sufficient intensity, it can thus give energy to the acoustic field faster than it is lost and fulfill a threshold condition which allows the acoustic energy to build up steadily

For the very high acoustic frequencies (10^9–10^{10} c p s) implied by Equation (13) when θ is not very small, the losses are usually so large that interesting amplification cannot be achieved with ordinary light But with laser beams of hundreds of megawatts per square centimeter, it is quite feasible to produce an intense build-up of acoustic waves by this process of stimulated Brillouin scattering [61, 62]—so intense, in fact, that the acoustic energy can crack glass or quartz This gives a method of producing and studying the behaviour of very high frequency acoustic waves in almost any material which will transmit light—a possibility which was previously not so clearly available

Brillouin scattering by spontaneous emission has been studied for some time But the intense monochromatic light of lasers allows now much greater precision in work with this technique [52], and it too is yielding

interesting information on the propagation of hypersonic waves in materials

There is no firm limit to the acoustic frequencies which can be produced by stimulated emission, even though Equation (13) indicates a kind of limit for $\theta = \pi$ of $2\nu_0 v/c$ But in the optical branch of acoustic waves, the phase velocity v can be very high In fact, stimulated Raman scattering, or the Raman maser mentioned briefly above, represents excitation of the optical branches of acoustic spectra, and generates coherent molecular oscillations Quantum-electronic techniques can thus allow interesting new ways to generate and explore most of the acoustic spectrum as well as much of the electromagnetic domain

Concluding Remarks

In a few years this brief report will no longer be of much interest because it will be outdated and superseded, except for some matters of general principle or of historical interest But it will happily be replaced by further striking progress and improved results We can look forward to another decade of rapid development in the field of quantum electronics—new devices and unsuspected facets of the field, improved range and performance of masers, and extensive application to science and to technology It seems about time now for masers and lasers to become every-day tools of science, and for the exploratory work which has demonstrated so many new possibilities to be increasingly replaced by much more finished, more systematic, and more penetrating applications It is this stage of quantum electronics which should yield the real benefits made available by the new methods of dealing with radiation

References

1 J R Pierce *Physics Today* **3**(11) 24 (1950)
2 N F Ramsey *Phys Rev* **103**, 20 (1956)
3 W E Lamb and R C Retherford, *Phys Rev* **79**, 549 (1950)
4 E M Purcell and R V Pound *Phys Rev* **81**, 279 (1951)
5 Bloch Hansen and Packard, *Phys Rev* **70**, 474 (1946)
6 N G Basov and A M Prochorov *J Exp Theor Phys USSR* **27**, 431 (1954)
7 J Weber *IEEE Trans Electron Devices* **3**, 1 (1953)
8 Columbia Radiation Lab Quarterly Progress Report Dec 1951 C H Townes *J Inst Elec Comm Eng Jpn* **36**, 650 (1953) (in Japanese)
9 Gordon Zeiger and Townes *Phys Rev* **95**, 282 (1954)
10 Gordon Zeiger and Townes *Phys Rev* **99**, 1264 (1955)
11 R Serber and C H Townes *Quantum Electronics* C H Townes (ed) p 233 Columbia Univ Press New York (1960) see also H Friedburg *ibid* p 227

12 Shimoda, Takahasi, and Townes, *J Phys Soc Jpn* **12**, 686 (1957), M W Muller, *Phys Rev* **106**, 8 (1957) M W P Strandberg, *Phys Rev* **106**, 617 (1957), R. V Pound, *Ann Phys* (*N Y*) **1**, 24 (1957)
13 C H Townes, *Advances in Quantum Electronics*, J R Singer (ed) p 3, Columbia Univ Press, New York (1961)
14 Shimoda, Wang and Townes, *Phys Rev* **102**, 1308 (1956)
15 A L Schawlow and C H Townes, *Phys Rev* **112**, 1940 (1958)
16 A correction of a factor of 2 to 8 given by references 10 and 15 is demonstrated by P Grivet and A Blaquiere Proceedings of Symposium on Optical Masers, Polytechnic Inst of Brooklyn April 1963
17 Alsop, Giordmaine, Townes, and Wang, *Phys Rev* **107**, 1450 (1957), J P Gordon and L D White, *Phys Rev* **107**, 1728 (1957)
18 J De Prins and P Kartaschoff, *Proceedings of Course XVII, Enrico Fermi International School of Physics* p 88, Academic Press New York (1962)
19 Goldenberg, Kleppner and Ramsey, *Phys Rev Lett* **5**, 361 (1960), *Phys Rev* **123**, 530 (1961)
20 Combrisson Honig and Townes, *C R Acad Sci* **242**, 2451 (1956), Feher, Gordon, Buehler Gere and Thurmond, *Phys Rev* **109**, 221 (1958)
21 N Bloembergen, *Phys Rev* **104**, 324 (1956)
22 N G Basov and A M Prochorov *J Exp Theor Phys USSR* **28**, 249 (1955)
23 Scovil Feher and Seidel *Phys Rev* **105**, 762 (1957)
24 Makhov, Kikuchi Lambe and Terhune *Phys Rev* **109**, 1399 (1958)
25 A E Siegman *Microwave Solid State Masers* McGraw-Hill New York (1964)
26 Alsop Giordmaine Mayer and Townes, *Astron J* **63**, 301 (1958), *Proc IEEE* **47**, 1062 (1959)
27 E Epstein *Astron J* **69**, 490 (1964)
28 R W DeGrasse and H E D Scovil *J Appl Phys* **31**, 443 (1960)
29 DeGrasse Hogg, Ohm, and Scovil, *J Appl Phys* **30**, 2013 (1959)
30 A L Schawlow and C H Townes *Phys Rev* **112**, 1940 (1958)
31 A G Fox and T Li *Bell Syst Tech J* **40**, 453 (1961)
32 Jacobs, Gould and Rabinowitz, *Phys Rev Lett* **7**, 415 (1961)
33 T H Maiman, *Nature* **187**, 493 (1960), Maiman, Hoskins D'Haenens, Asawa, and Evtuhov *Phys Rev* **123**, 1151 (1961)
34 Collins, Nelson, Schawlow, Bond Garrett and Kaiser *Phys Rev Lett* **5**, 303 (1960)
35 Javan Bennett and Herriott *Phys Rev Lett* **6**, 106 (1961)
36 A Javan, *Phys Rev Lett* **3**, 87 (1959)
37 R J Keyes and T M Quist *Proc IEEE* **50**, 1822 (1962)
38 Hall Fenner, Kingsley Soltys and Carlson, *Phys Rev Lett* **9**, 366 (1962)
39 Nathan, Dumke, Burns, Dill, and Lasher *Appl Phys Lett* **1**, 62 (1962)
40 Quist, Rediker, Keyes, Krag, Lax McWhorter, and Zeiger, *Appl Phys Lett* **1**, 91 (1962)
41 Basov Bogdankevich and Devyatkov *Dokl Akad Nauk SSSR* **155**, 783 (1964) Proceedings Symposium on Radiative Recombination in Semiconductors Paris 1964, Dunod (to be published) C Benoit à la Guillaume and T M De Bever *ibid*
42 A Javan, *J Phys Radium* **19**, 806 (1958), *Bull Am Phys Soc* **3**, 213 (1958), *Proceedings of Course XXXI* (1963) *Enrico Fermi International School of Physics*, Academic Press New York (1964)
43 E J Woodbury and W K Ng *Proc IEEE* **50**, 2367 (1962) Eckhardt, Hellwarth McClung Schwarz Weiner, and Woodbury, *Phys Rev Lett* **9**, 455 (1962)
44 R W Terhune, *Solid State Design* **4**, 38 (1963), Minck, Terhune and Rado *Appl*

Phys Lett **3**, 181 (1963) see also B Stoicheff *Proceedings of Course XXXI* (1963) *Enrico Fermi International School of Physics* Academic Press (1965)
45 R W Hellwarth, *Phys Rev* **130**, 1850 (1963) H J Zeiger and P E Tannenwald *Proc Third Quantum Elec Conf* P Grivet and N Bloembergen (eds) p 1589, Dunod Paris (1964) Garmire Pandarese and Townes, *Phys Rev Lett* **11**, 160 (1963)
46 E L Tolnas, *Bull Am Phys Soc* **5**, 342 (1960)
47 Jaseja, Javan and Townes *Phys Lett* **10**, 165 (1963)
48 D Herriott, *J Opt Soc Am* **52**, 31 (1962)
49 J P Cedarholm and C H Townes, *Nature* **184**, 1350 (1959)
50 Jaseja Javan Murray and Townes *Phys Rev* **133** A1221 (1964)
51 A Szoke and A Javan *Phys Rev Letts* **10**, 521 (1963)
52 R Chiao and B Stoicheff *J Opt Soc Am* **54**, 1286 (1964) Benedek Lastovka, Fritsch and Greytak *J Opt Soc Am* **54**, 1284 (1964) Masch Starunov, and Fabelinskii *J Exp Theor Phys USSR* **47**, 783 (1964)
53 C A Akhmanov and R B Khokhlov *Problems in Nonlinear Optics 1962–1963* Acad Sci U S S R Moscow (1964)
54 N Bloembergen *Nonlinear Optics* W Benjamin, New York (1965)
55 I Abella *Phys Rev Lett* **9**, 453 (1962)
56 J A Giordmaine and J A Howe *Phys Rev Lett* **11**, 207 (1963)
57 Franken Hill Peters and Weinreich *Phys Rev Lett* **7**, 118 (1961) P A Franken and J F Ward *Rev Mod Phys* **35**, 23 (1963)
58 J A Giordmaine *Phys Rev Lett* **8**, 19 (1962) Terhune Maker and Savage *Appl Phys Lett* **2**, 54 (1963)
59 J Ducuing and N Bloembergen *Phys Rev Lett* **10**, 474 (1963)
60 Armstrong Bloembergen Ducuing and Pershan *Phys Rev* **127**, 1918 (1962) N Bloembergen and P S Pershan *Phys Rev* **128**, 606 (1962)
61 Chiao Garmire and Townes *Proceedings of Course XXXI* (1963) *Enrico Fermi International School of Physics* Academic Press New York (1965)
62 Chiao, Townes and Stoicheff *Phys Rev Lett* **12**, 592 (1964)
63 C H Townes *Quantum Electronics* C H Townes (ed) p 402 Columbia Univ Press New York (1960) C Kittel *Phys Rev Lett* **6**, 449 (1961)
64 E B Tucker *Phys Rev Lett* **6**, 547 (1961)
65 L Brillouin *Ann Phys Paris* **17**, 88 (1922) M Born and K Huang *Dynamical Theory of Crystal Lattices* p 373 Oxford Univ Press London (1954)

Quantum Electronics

A. M. Prochorov

One may assume as generally accepted that quantum electronics started to exist at the end of 1954—beginning of 1955 [1, 2]. Just by that time theoretical grounds had been created, and the first device—a molecular oscillator—had been designed, and constructed. A basis for quantum electronics as a whole is the phenomenon of an induced radiation, predicted by A. Einstein in 1917. However, quantum electronics was developed considerably later. And it is quite natural to ask the questions why did it happen so? What reasons put obstacles for the creation of quantum devices considerably earlier, for instance, in the period of 1930–1940?

In order to try to answer these questions I should like to say some words about the bases on which quantum electronics is founded.

As I have already noted, the phenomenon of an induced radiation was predicted by Einstein. It is well-known that an atom being in an excited state may give off its energy in the form of emission of radiation (quantum) in two ways. The first way is a spontaneous emission of radiation, i.e. when an atom emits energy without any external causation. All usual light sources (filament lamps, gas discharge tubes, etc.) produce light by way of such spontaneous radiation. It means that scientists engaged in the field of optical spectroscopy were well acquainted with that type of emission many years ago.

The second way for an atom to give off its energy is through stimulated emission of radiation. That phenomenon was noted by Einstein to be necessary in order to describe thermodynamic equilibrium between an electromagnetic field and atoms. The phenomenon of stimulated emission occurs when an excited atom emits due to interaction with an external field (quantum). Then two quanta are involved: one is the external one, the other is emitted by the atom itself. Those two quanta

are indistinguishable, i.e. their frequency and directivity coincide. This very significant characteristic of an induced radiation (which was, apparently, first pointed out by Dirac in 1927) made it possible to build quantum electronic devices.

In order to observe a stimulated emission, it is necessary, firstly, to have excited atoms and, secondly, that the probability of an induced radiation must be greater than that of a spontaneous emission. If atoms are in a thermal equilibrium, optical levels are not populated. If the atoms become excited they make a transition to the lower level due to spontaneous emission. This happens because the probability of a stimulated emission radiation is small at usual densities of the light energy. Therefore scientists engaged in the field of spectroscopy did not take into account the stimulated radiation and some of them, apparently, considered that phenomenon as a "Kunststuck" of a theorist necessary only for the theory.

It is absolutely clear that if all atoms are in an excited state, such a system of atoms will amplify the radiation, and many scientists understood this already before 1940, but none of them pointed to the possibility of creating light oscillators in this way. It may seem strange because, in principle, optical quantum oscillators (lasers) could have been made even before 1940. But definite fundamental results were necessary. They appeared after the Second World War, when radiospectroscopy started to develop rapidly. And just the scientists engaged in the field of radiospectroscopy laid down foundations for quantum electronics [1, 2].

How should one explain this? There were some favourable circumstances which had not been available to the scientists working in the field of optical spectroscopy.

First of all, since for systems in a thermal equilibrium the excited levels in the radio range, contrary to the optical ones, may have a large population and of course one should then take into account induced radiation. Indeed, if the concentration of particles on the lower level equals n_1, and on the excited level n_2, one may write down a net absorption coefficient for an electromagnetic wave in the form

$$\alpha = \frac{1}{v} h\nu(n_1 B_{12} - n_2 B_{21}) \tag{1}$$

The value of B_{12} characterizes the probability of an absorption act, and B_{21} characterizes the probability of an induced radiation act. If the levels are not degenerate $B_{12} = B_{21}$, then (1) will take the form

$$\alpha = \frac{1}{v} h\nu(n_1 - n_2) B_{12} \tag{2}$$

For a frequency as in the optical range, under usual conditions of thermal equilibrium, one may put with a high accuracy n_2 equal to zero, and then the absorption coefficient will become

$$\alpha = \frac{1}{v} h\nu n_1 B_{12} \tag{3}$$

Therefore, for the optical range the absorption coefficient depends only on the population of the lower level For as in the ratio range, as a rule, $h\nu \ll kT$ In that case

$$n_2 = n_1 e^{-h\nu/kT} \approx n_1\left(1 - \frac{h\nu}{kT}\right)$$

Then the value of a will be

$$\alpha = \frac{1}{v} h\nu n_1 B_{12} \frac{h\nu}{kT} \tag{4}$$

As is seen from (4), due to stimulated emission, the value of the absorption coefficient becomes reduced by a factor $kT/h\nu$ compared to what it would be without the presence of induced emission Therefore, all scientists engaged in radiospectroscopy have to take into account the effect of induced radiation Moreover, for increasing an absorption coefficient one has to lower the temperature in order to decrease the population of the upper level and to weaken, in this way, the influence of stimulated radiation It follows from (2) that for systems that are not in thermal equilibrium, but have $n_2 > n_1$, the net absorption coefficient becomes negative, i e such a system will amplify radiation In principle, such systems were known to physicists a long time ago for the radio range If we pass molecular or atomic beams through inhomogeneous magnetic or electric fields, we can separate out molecules in a definite state In particular, one may obtain molecular beams containing molecules in the upper state only Actually physicists engaged in the field of microwave radiospectroscopy started to think about application of molecular beams for increasing the resolving power of radio spectroscopes In order to gain a maximum absorption in such beams one must have molecules either in the lower state only or in the upper state only i e one must separate them using inhomogeneous electric or magnetic fields If molecules are in the upper state, they will amplify a radiation

As is well-known from radio engineering, any system able to amplify can be made to oscillate For this purpose a feedback coupling is necessary A theory for ordinary tube oscillators is well developed in the radio range For description of those oscillators, the idea of a negative resistance or conductance is introduced, i e an element in which so-called

negative losses take place In the case of a quantum oscillator the medium with a negative absorption factor is that "element " Therefore the condition of self-excitation for the quantum oscillator should be written in the similar way as for a tube oscillator According to the analogy with usual tube oscillators, it is quite natural to expect that for a quantum oscillator the oscillations will also be quite monochromatic

Finally, the resonator system is a very significant element of a quantum oscillator (maser or laser) as well as in any other oscillator with sinusoidal oscillations However, resonator systems were well worked out for the radio range, and just those resonators operating in the radio range were used for masers Thus, a very important element—a cavity—was also well-known to the scientists engaged in radiospectroscopy Therefore all elements of masers really existed separately but it was necessary to do a very important step of synthesis in order to construct the maser First, two papers—one of which was published in the USSR and the other in the USA—appeared independently, and they both were connected directly with the construction of radio spectroscopes with a high resolving power, using molecular beams As is easily seen from the aforesaid, this result is quite natural

Those two papers initiated the development of quantum electronics, and the first successes in this new field of physics stimulated its further progress Already in 1955, there was proposed a new method—the method of pumping for gaining a negative absorption [3] That method was further developed and applied for the construction of new types of quantum devices In particular, the method of pumping was developed and applied for designing and building quantum amplifiers for the radio range on the basis of an electronic paramagnetic resonance [4, 5] Quantum devices according to the suggestion of Professor Townes were called masers One might think that after the successful construction of masers in the radio range, there would soon be made quantum oscillators (lasers) in the optical range as well However, this did not occur Those oscillators were constructed only 5–6 years later What caused such a delay?

There were two difficulties One of them was as follows at that time no resonators for the optical wavelength range were available The second difficulty was that no methods were immediately available for gaining an inverse population in the optical wavelength range

Let us consider first the question of resonators It is well-known that radio engineering started its development from the region of long waves where resonators were used in the form of self-inductance coils combined with condensers In that case the size of the resonator is much less than one wavelength With development towards short waves the cavity resonators were used They are closed volume cavities The size of

those cavities was comparable with a wavelength It is quite clear that with the help of such cavities it is impossible to advance into the region of very short waves In particular, it would be impossible to reach the optical range

In 1958 there was proposed the so-called open type of cavities for masers and lasers in the region of very short waves [6, 7] Practically speaking this is Fabry–Perot's interferometer, however, a "radio engineering" approach made it possible to suggest using such a system as a resonator Afterwards, spherical mirrors were used together with plane mirrors The size of these resonators is much more than the wavelength

At present open cavities are widely utilized for lasers

There were also systems suggested for the production of a negative absorption in the submillimeter (far infrared) wavelength range [6] the infrared and optical wave ranges [7, 8, 9, 10] Those works stimulated a further advancement in the region of shorter waves and in particular into the optical range However, the first quantum optical oscillator was made as late as 1960 [11] It was a ruby laser After carrying out investigations in the optical range, many scientists started to think about further extension into the X-ray field In that wavelength range the same difficulties arise as in the optical wavelength range It was necessary to suggest new types of resonators and to find also the proper system that would produce negative absorption As is known X-ray quantum oscillators have not yet been constructed We have also considered this problem† and we have found that there are essential difficulties

Indeed in the X-ray region the lifetime of an excited level state is small and one may assume that the line width is determined by that lifetime only Then the absorption coefficient may be written in a very simple form

$$\alpha = \frac{\lambda^2}{4\pi}(n_1 - n_2) \qquad (5)$$

where λ is the wavelength and n_1 and n_2 are the densities of particles in the lower and upper level states respectively As is seen from (5) the absorption coefficient decreases sharply as the wavelength becomes shorter This is an extremely unpleasant circumstance Indeed for the laser operation the value of α should be of the order of one inverse cm If $\lambda = 1$ Å, the density of particles on the upper level must be not less than 10^{17} cm^{-3} The lifetime in the upper level τ is of the order of 10^{-16} sec Therefore, 10^{33} particles/cm^3 per second must be excited In order to fulfill this condition one has to overcome essential experimental difficulties

† This problem was also considered in Reference [1]

Nevertheless, the successes of quantum electronics are enormous even without the construction of lasers in the X-ray region

At present the range in which lasers and masers operate is extremely wide. Recently a far infrared range had not been available but now investigations in this region are carried out with great success In practice, with the help of masers and lasers one may produce emission from the lowest radio frequencies to the ultraviolet region

Operation of all masers and lasers is based on the fact that in media with a negative absorption the processes of induced emission dominate due to a large field intensity over spontaneous or non-radiative transitions Moreover, at present, for instance, one may produce with the help of a ruby laser such radiation energy densities at which the probability of multiquantum processes becomes comparable with the probability of one quantum process or even exceeds it This is a new qualitative jump which leads to interesting results of several kinds

First of all one may estimate [12] the maximum power which a ruby laser is able to give per one cm^2 That power equals 10^{11} watt/cm^2, that is one hundred gigawatts/cm^2 At that power the probability of a simultaneous absorption of three quanta of red light with transition of an electron to the conduction band is so great that a further growth of the field stops For three-quantum processes the losses grow in proportion to the cube of the energy density, i e a very strong dependence on the field takes place

Large electric fields available in a laser beam may carry out ionization and dissociation of molecules and breakdown in a solid as well

Multiquantum processes do not always have a bad effect (for instance, restriction of the maximum density given by laser) but they open up new possibilities for a further development of quantum electronics This interesting and principally new direction is connected with the construction of lasers which utilize two-quantum transitions It was pointed out in 1963 in the USSR [13] and somewhat later but independently in the USA [14, 15] that construction of these oscillators should be possible The idea of this laser is that if there is an inverse population between two levels with the energy difference $E_2 - E_1 = h\nu$, generation of two frequencies ν_1 and ν_2 is possible in such a way that

$$\nu = \nu_1 + \nu_2 \tag{6}$$

In particular, frequencies ν_1 and ν_2 may coincide However, frequencies ν_1 and ν_2 may have any value as long as only condition (6) is fulfilled

Operation of such an oscillator, as was mentioned above, is connected with two-quantum transitions, the probability of which is rather great, if the field density is considerable For self-excitation of these oscillators it is necessary to have another oscillator of a sufficiently large

initial energy density with frequencies ν_1 or ν_2, and one may remove the external field only after self-excitation of the two-quantum oscillator Such two-quantum oscillators have two possibilities

1) Faster growth of the field density than in the case of usual lasers
2) Possibility of producing any frequency within the frameworks of the relation (6)

Construction of an oscillator for any given radiation frequency will greatly extend the region of application of lasers It is clear that if we make a laser with a sweep frequency, we apparently shall be able to influence a molecule in such a way that definite bonds will be excited and, thus, chemical reactions will take place in certain directions

However, this problem will not be simple even after design of the appropriate lasers But one thing is clear the problem is extremely interesting and perhaps its solution will be able to make a revolution in a series of branches of chemical industry

References

1 N G Basov and A M Prochorov, *Zh Eksp Teor Fiz* **27**, 431 (1954)
2 J P Gordon, H J Zeiger, and C H Townes, *Phys Rev* **95**, 282 (1954)
3 N G Basov and A M Prochorov, *Zh Eksp Teor Fiz* **28**, 249 (1955)
4 N Bloembergen, Phys *Rev* **104**, 324 (1956)
5 H E D Scovil, G Feiher, and H Seidel, *Phys Rev* **105**, 762 (1957)
6 A M Prochorov *Zh Eksp Teor Fiz* **34**, 1658 (1958)
7 A L Schawlow and C H Townes, *Phys Rev* **112**, 1940 (1958)
8 N G Basov, B M Vul, Yu M Popov, *Zh Eksp Teor Fiz* **37**, 587 (1959)
9 A Javan *Phys Rev Lett* **3**, 87 (1959)
10 F A Butaeva and V A Fabrikant Media with negative coefficients of absorption Investigations in experimental and theoretical physics *Collection of Works in Memory of G S Landsberg*, p 62 published by the Academy of Sciences of the USSR, Moscow, (1959)
11 T H Maiman, *Brit Comm Electr* **1**, 674 (1960)
12 F V Bunkin and A M Prochorov, *Zh Eksp Teor Riz*, in press
13 A M Prochorov and A C Selivanenko, Russian Patent Application No 872303, with priority from December 24, 1963
14 P P Sorokin, and N Braslau, *IBM J* **8**, 177 (1964)
15 R Z Gorwin, *IBM J* **8**, 338, (1964)

Nobel Lecture, December 11, 1964

Semiconductor Lasers

Nicolai G. Basov

In modern physics, and perhaps this was true earlier, there are two different trends One group of physicists has the aim of investigating new regularities and solving existing contradictions They believe the result of their work to be a theory, in particular, the creation of the mathematical apparatus of modern physics As a by-product there appear new principles for constructing devices, physical devices

The other group, on the contrary, seeks to create physical devices using new physical principles They try to avoid the inevitable difficulties and contradictions on the way to achieving that purpose This group considers various hypotheses and theories to be the by-product of their activity

Both groups have made outstanding achievements Each group creates a nutrient medium for the other and therefore they are unable to exist without one another, although, their attitude towards each other is often rather critical The first group calls the second "inventors," while the second group accuses the first of abstractness or sometimes of aimlessness One may think at first sight that we are speaking about experimenters and theoreticians However, this is not so, because both groups include these two kinds of physicists

At present this division into two groups has become so pronounced that one may easily attribute whole branches of science to the first or to the second group, although there are some fields of physics where both groups work together

Included in the first group are most research workers in such fields as quantum electrodynamics, the theory of elementary particles, many branches of nuclear physics, gravitation, cosmology, and solid-state physics

Striking examples of the second group are physicists engaged in

thermonuclear research, and in the fields of quantum and semiconductor electronics

Despite the fact that the second group of physicists strives to create a physical device, their work is usually characterized by a preliminary theoretical analysis Thus, in quantum electronics, there was predicted theoretically the possibility of creating quantum oscillators in general, also, there were predicted the high monochromaticity and stability of the frequency of masers, the high sensitivity of quantum amplifiers, and there was investigated the possibility of the creation of various types of lasers

This lecture is devoted to the youngest branch of quantum electronics—semiconductor lasers, which was created only two years ago, although a theoretical analysis started already in 1957 preceded the creation of lasers [1]

However, before starting to discuss the principles of operation of semiconductor lasers we would like to make some remarks on the theoretical "by-products" of quantum electronics There are many of them but we shall consider only three

1 The creation of quantum frequency oscillators of high stability and the transition to atomic standards of time made it possible to raise the question of solving the problem of the properties of atomic time

Dicke [2] in his paper at the first conference on quantum electronics pointed out the possibility of an experimental check of the hypothesis on the variation of fundamental physical constants with time on the basis of studying changes in frequencies of different quantum standards with time There arises the question about the maximum accuracy of atomic and molecular clocks depending on the nature of quantum of emission, especially about the accuracy of the measurement of short time intervals

2 Due to quantum electronics there was started an intensive investigation of a new "super non-equilibrium state of matter"—the state with negative temperature, which in its extreme state of negative zero is close in its properties to the absolute ordering intrinsic for the temperature of absolute zero It is just this property of high ordering of a system with negative temperature which makes it possible to produce high-coherent emission in quantum oscillators, to produce high sensitive quantum amplifiers, and to separate the energy stored in the state with negative temperature in a very short time, of the order of the reciprocal of the emission frequency

3 Quantum electronics gives examples of systems in which there occurs radiation with a very small value of entropy For instance, spontaneous low temperature radiation from flash tubes, distributed through very large number of degrees of freedom is converted with the help of a system in a state of negative temperature (quantum oscillators) into

high-coherent laser emission, the temperature of which in present experiments already attains a value of 10^{20} degrees

Apparently, the regularities established by quantum electronics for radiation may be generalized for other natural phenomena The possibility of obtaining a high degree of organization with the help of feedback systems may be of interest for chemical and biological research, and for cosmology The question arises as to whether or not the maser principle is used in Nature

We believe that the above questions need attention from physicists of the first group, because these questions go beyond the limits of the theory of oscillations, the theory of radiation and usual optics which form the basis of modern quantum electronics

I. Conditions for the Production of Negative Temperature in Semiconductors

Investigations of semiconductor quantum oscillators were a direct continuation of research on molecular oscillators and paramagnetic amplifiers One should note that at the beginning of research on semiconductor lasers, due to investigations in the field of semiconductor electronics, there became known the physical characteristics of semiconductors, which were essential in the development and practical realization of lasers, such as, optical and electric characteristics, structure of energy bands, and relaxation time

Various pure and alloyed semiconductors were made, and the technique of measurements of their various properties and the technology of making *p–n* junctions were worked out All of this considerably simplified investigations of semiconductor lasers Semiconductors were very intriguing because of the possibility of using them for making oscillators with a frequency range from the far infrared region to the optical or even to the ultraviolet range, as well as because of the variety of methods by means of which states with negative temperatures may be obtained within them and because of their large factor of absorption (amplification) As the following studies have shown, semiconductor lasers may have extremely high efficiency, in some cases approximating 100 percent

In contrast to an isolated atom, in semiconductors there do not exist separate energy levels, but rather there exist groups of energy levels arranged very close to one another, which are called bands (Fig 1) The upper group of levels, called the conduction band, and a lower group of excited levels, called the valent band, are divided by a band of forbidden energy (Fig 1)

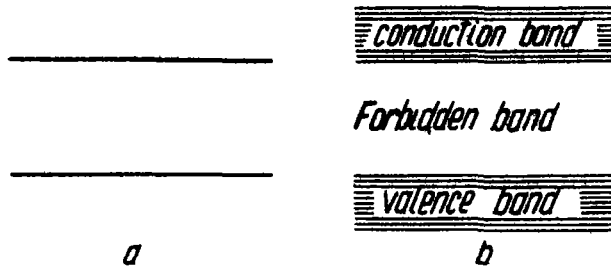



Fig. 1 Energy-level diagram (a) For atom with two energy levels (b) For semiconductor

The distribution of electrons in energy levels is described by the Fermi function each level is occupied by two electrons, the electrons being distributed in the energy range of the order of the energy of kT thermal motion, and, the probability of finding an electron beyond the kT interval sharply decreases when the energy level increases If the energy of thermal motion is significantly less than the energy difference between the conduction and valent bands, then practically all electrons will be found in the valent band, filling its levels, while practically all levels of the conduction band remain free (Fig 2a) In such a state the semiconductor cannot conduct electric current and becomes an insulator, since the electric field applied to the semiconductor is unable to change the motion of the electrons in the valent band (all energy levels are occupied)

If the energy of thermal motion is sufficient, then a part of the electrons are thrown into the conduction band Such a system may serve as a conductor of electric current Current is able to flow both due to variation of the electrons' energy under the action of the external field, as well as due to changes in the electron distribution within the valence and

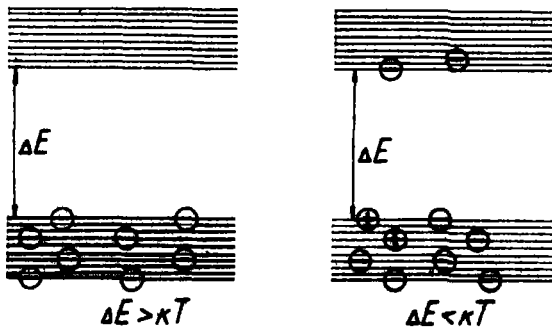



Fig 2 Distribution of the electrons on energy levels

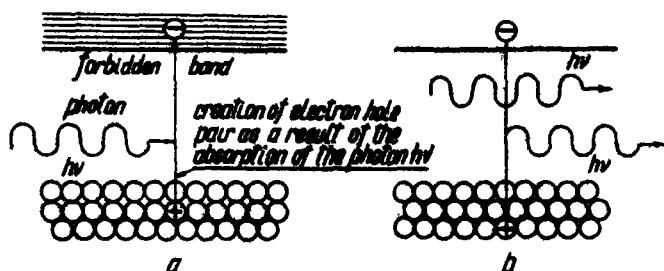



Fig 3 Processes of the interaction with light (a) Resonance absorption (b) Stimulated emission

conduction bands Current within the valent band behaves as if those places free from electrons (holes) moved in a direction opposite to that of the electrons A vacant place or "hole" is entirely equivalent to a positively charged particle (Fig 2b)

During interaction with light, a semiconductor, similar to an isolated atom, may undergo three processes

1 A light quantum may be absorbed by the semiconductor, and, in this case an electron–hole pair is produced The difference in energy between the electron and the hole is equal to the quantum energy This process is connected with the decrease in energy of the electromagnetic field and is called *resonance absorption* (Fig 3a)

2 Under the influence of a quantum, an electron may be transferred from the conduction band to a vacant place (hole) on the valent band Such a transfer will be accompanied by the emission of a light quantum identical in frequency, direction of propagation, and polarization to the quantum which produced the emission This process is connected with an increase of the field energy and is called *stimulated emission* (Fig 3b) We recall that stimulated emission was discovered by A Einstein in 1917 during an investigation of thermodynamical equilibrium between the radiation field and atoms

3 Besides resonance absorption and simulated emission, a third process may take place—*spontaneous emission* An electron may move over to a vacant place-hole (recombine with the hole) in the absence of any radiation quanta

Since the probabilities of stimulated radiation and resonance absorption are exactly equal to one another, a semiconductor in an equilibrium state at any temperature may only absorb light quanta, because the probability of finding electrons at high levels decreases as the energy increases In order to make the semiconductor amplify electromagnetic radiation, one must disturb the equilibrium of the distribution of electrons

within the levels and artificially produce a distribution where the probability of finding electrons on higher energy levels is greater than that of finding them on the lower levels [1, 3] It is very difficult to disturb the distribution inside a band because of the strong interaction between the electrons and the lattice of the semiconductor it is restored in 10^{-10}–10^{-12} sec It is much simpler to disturb the equilibrium between the bands, since the lifetime of electrons and holes is considerably greater in the bands It depends on the semiconductor material and lies in the interval 10^{-3}–10^{-9} sec

Due to the fact that electrons and holes move in semiconductors, in addition to the law of the conservation of energy, the law of the conservation of momentum should be fulfilled during emission Since the photon impulse is extremely small, the law of the conservation of momentum, approximately speaking, requires that the electrons and holes must have the same velocity during the emission (or absorption) of a light quantum Figure 4 shows graphically the dependence of energy on momentum There are two types of semiconductors For one group of semiconductors, the minimum of electron energy in the conduction band is exactly equal to the maximum of hole energy in the valent band In such semiconductors there may take place so called "direct transitions " An electron having minimum energy may recombine with a hole having maximum energy For another group of semiconductors, the minimum energy in the conduction band does not coincide with the maximum energy in the valent band In this case the process of emission or absorption of a light quantum should be accompanied by a change in the amplitude of the oscillatory state of the crystal lattice, that is by the emission or absorption of a phonon which should compensate for the change in momentum Such processes are called indirect transitions The

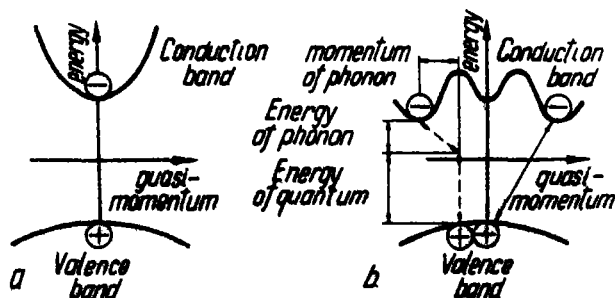



Fig. 4 Diagram of the electron–hole energy dependence on the quasimomentum (a) Direct transitions (b) Indirect transitions

probability of indirect transitions is usually less than that of direct transitions

In order to make a semiconductor amplify incident radiation under inter-band transitions, one should distinguish two cases

(a) In the Case of Direct Transitions

It is necessary to fill more than half of the levels in the band of the order of kT near the band's edge with electrons and holes Such states, both for atoms and molecules, came to be called states with inverse populations, or states with negative temperature The distribution of electrons when all levels in the kT zone of the conduction band are occupied by electrons, and in the valent band—by holes, corresponds to the temperature minus zero degrees In this state (in contrast to the state of plus zero degrees), the semiconductor is only able to emit (stimulated and spontaneous) light quanta and is unable to absorb emissions

The state of a semiconductor when most levels in a certain energy band are occupied by electrons or holes was named the *degenerated state*

Thus, for the creation of negative temperature there must occur degeneration of electrons and holes in the semiconductor With a given number of electrons and holes it is always possible to produce degeneration by means of lowering the semiconductor's temperature, since, as the temperature decreases the energy band width occupied by the electrons also decreases At the temperature of liquid nitrogen for degeneration to take place it is necessary to have an electron concentration of 10^{17}–10^{18} l/cm^3 [3]

(b) In the Case of Indirect Transitions

Degeneration is not necessary for the creation of negative temperature This is connected with the fact that when indirect transitions occur, the probability of quantum stimulated emission may not be equal to the probability of resonance absorption

Consider, for instance, an indirect transition in which a quantum and a phonon are emitted simultaneously The process of the simultaneous absorption of a quantum and a phonon is the inverse of that process

The probability of absorption is proportional to the number of phonons in the crystal lattice The number of phonons decreases with a lowering of temperature At low temperature phonons are absent Therefore, by means of lowering the temperature of the sample one may make the probability of emission much greater than the probability of absorption This means that with indirect transitions negative temperature may

be attained with a considerably lower concentration of electrons and holes [4]

One should note that the absorption and emission of quanta during transitions within a band also takes place due to indirect transitions When negative temperature is created between bands, the distribution of electrons (and holes) within a band corresponds to a positive temperature and leads to the absorption of emission

In the case of direct transitions, when the probability of interband transitions is much greater than that of innerband transitions, one may neglect the innerband transitions, that is, states with negative temperature can be used for the amplification of emissions

In the case of indirect transitions for amplification to take place it is not sufficient to attain negative temperature It is necessary that the probability of interband transitions be greater than that of innerband transitions The necessity of fulfilling this condition makes it difficult to utilize indirect transitions According to Dumke's estimate, this condition cannot be fulfilled for germanium [5] However, it may be fulfilled for other semiconductors [6]

In a number of cases in semiconductors, an electron and a hole form an interconnected state something like an atom–exiton The exitons may recombine, producing an emission They may be also used for the manufacture of quantum amplifiers, but we shall not consider this in detail

We have studied conditions for the production of negative temperature in semiconductors possessing an ideal lattice In a nonideal crystal there occur additional energy levels connected with various disturbances in the crystalline lattice (impurities, vacancies, dislocations, etc) As a rule, these states are localized near the corresponding center (for instance, near an impurity atom) and in this they differ from those states in the valent and conduction bands which belong to the crystal as a whole

In an ideal crystal the number of electrons in the conduction band is exactly equal to the number of holes in the valent band However, in an actual crystal the number of current carriers—electrons and holes—is determined, mainly, by the existence of impurities (Fig 5)

There are two kinds of impurities one type has energy levels arranged near the conduction band and creates excess electrons due to thermal ionization These are called "donor" impurities Other impurities having energy levels near the valent band are capable of removing electrons from the valent band and thus producing an excess number of holes in it These impurities are called "acceptors "

One should note that a semiconductor with an equal number of donor and acceptor impurities behaves as if it were a pure semiconductor,

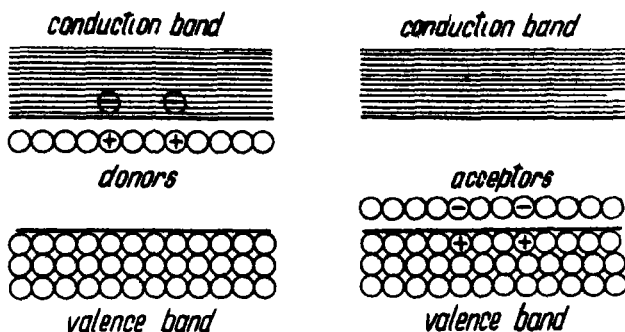



Fig 5 Donor and acceptor levels

since the holes produced by acceptors recombine with the electrons produced by donors

In a number of cases, transitions of electrons between bands, between impurity atoms, or between levels may also be accompanied by emission of photons One may likewise use these transitions for the creation of negative temperature However, because of time limitations, we shall not discuss this question

II. Methods of Obtaining States with Negative Temperature in Semiconductors

(a) The Method of Optical Pumping

In the case of semiconductors one may utilize the "three-level" scheme [7] which has been used successfully for paramagnetic quantum amplifiers [8] and optical generators based on luminescent crystals and glasses [9] (Fig 6)

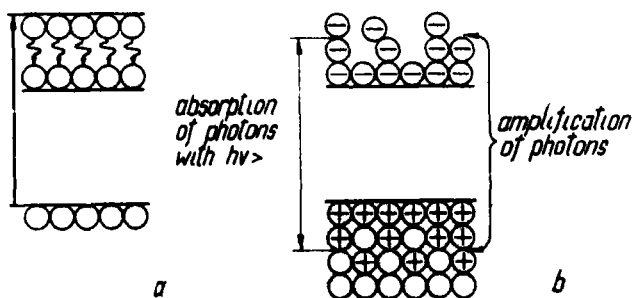



Fig 6 Optical pumping (a) Three-levels diagram for atoms (b) For semiconductors

Since the relaxation time of electrons and holes in the band lev
[10] is much less than the lifetime of electrons and holes in the corre
ponding bands, one may obtain an inverse population by means of optic
pumping

Semiconductors have a very large absorption index which sharp
increases as the radiation frequency increases Therefore, to obtain
inverse population in samples of relatively large thickness, it is reasonab
to use monochromatic radiation with a frequency close to that of t
interband transitions [11] In the case when the frequency of the exciti
radiation is greater than the width of the forbidden band, a state wi
negative temperature is produced in a narrow band several microns de
(on the order of the electrons' diffusion length) near the surface of t
sample As a source of radiation one may use the light from other typ
of lasers gas lasers, lasers based on luminescent crystals, or lasers bas
on p–n junctions [11]

(b) *The Excitation of Semiconductors by a Beam of Fast Electrons*

If a beam of fast electrons is directed onto the surface of a sem
conductor, the electrons easily penetrate into the semiconductor On the
way the electrons collide with the atoms of the crystal and crea
electron–hole pairs Calculations and experiments [12, 13] have show
that an amount of energy approximately three times greater than t
minimum energy difference between the bands is spent on the productio
of one electron–hole pair (Fig 7a) The electrons and holes obtained gi
their excess energy to the atoms of the lattice and accumulate in the leve
near the edges of the corresponding bands In this case a state wi
negative temperature may be created [14, 15] The higher the electron

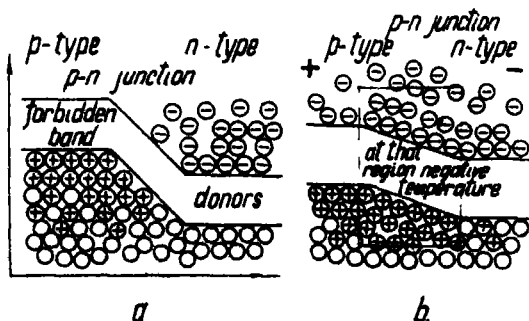



Fig 7 (a) p–n junctional equilibrium (b) p–n junction in the external electrical field

energy, the deeper they will penetrate However, there exists a certain threshold energy, beginning with which the electrons will produce defects in the crystal, that is, will destroy the crystalline lattice This threshold energy depends upon the binding energy of the atoms in the crystals and is usually equal to about several hundred keV Experiments have shown that electrons with energy in the range of 200–500 keV are not yet capable of noticeably harming the lattice

The current density of fast electrons at which negative temperature is produced strongly depends upon the lifetime of electrons and holes For semiconductors with a lifetime of 10^{-7} sec at the temperature of liquid nitrogen, the threshold of the current density has the order of one ampere per cm^2 Since in the presence of such large currents it is difficult to remove the energy released in the semiconductor, the impulse method of excitation with a short impulse duration is usually employed

(c) *The Injection of Electrons and Holes through p–n Junctions*

As was noted above, a specific characteristic of semiconductors is that their energy levels may be filled with electrons or holes by introducing into the crystals special types of impurity atoms However, the simultaneous introduction of donor and acceptor impurities does not result in the production of states with negative temperature Therefore, in order to obtain an inverse population one does as follows take two pieces of a semiconductor, inject donor impurities into one of them, and inject acceptor impurities into the other If one then connects one piece to the other, a *p–n* junction will be created On the boundary between the semiconductors there arises a potential difference which does not allow electrons to penetrate into the crystal having holes and likewise does not allow holes to penetrate into the crystal having electrons (Fig 7a) As was pointed out above, a large concentration of electrons and holes is necessary for the production of an inverse population (more than half of the levels in a certain energy band should be occupied), that is the semiconductor must have a large number of impurities

If one applies an external voltage to a *p–n* junction, removing the potential difference between the two pieces of the semiconductor, the equilibrium of the distribution of electrons will be disturbed, and current will flow through the semiconductor In this case electrons appear to flow into the region with a large concentration of holes, and holes—into the region with a large concentration of electrons An inverse population arises in a narrow region near the *p–n* junction at a distance of several microns Thus, there is obtained a layer of the semiconductor which is

able to amplify electromagnetic waves by means of the stimulated emission of quanta during the transition of electrons from the conduction band to the valent band [16] (Fig 7b)

Many methods for the production of p–n junctions were worked out during research on semiconductors At the present time two methods of making p–n junctions are used for the creation of lasers the diffusion method [17, 18] and the method of doping with different impurities during the process of growing a crystal [19]

III. Semiconductor Lasers

In order to carry out generation on the basis of systems with negative temperature, one must introduce feedback coupling into the system This feedback coupling is carried out with the aid of cavities The simplest type of cavity in the optical range is a cavity with plane parallel mirrors [20, 21] Light quanta reflecting from the mirrors will pass many times through the amplifying medium If a light quantum, before its absorption by the mirrors or inside the sample, has time to cause stimulated emission of more than one quantum (that is, if the condition of self-excitation is fulfilled in the system), that system will operate as a laser (Fig 8) If one maintains a certain negative temperature in the sample with the help of an external energy source, the number of quanta in the cavity will increase until the quantity of electrons excited per time unit becomes equal to the number of emitted quanta

It should be especially noted, that when a quantum system with feedback coupling operates as a laser, its emission has a very narrow frequency band This characteristic makes laser emission different from all other light sources filament lamps, luminescent lamps and light sources with very narrow atomic and molecular spectral lines

The monochromatic emission of lasers is a result of the properties of stimulated radiation the quantum frequency of the stimulated radiation

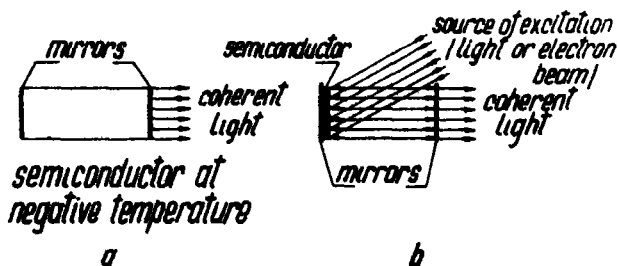



Fig 8 Diagram of semiconductor lasers (a) Usual (b) With radiative mirrors

equals the frequency of the quantum which produced the radiation The initial line width in semiconductors is usually about several hundred angstroms At the present time it has been shown that the line width in lasers which use a *p–n* junction in GaAs is less than fifty megacycles [22, 24] The minimum value of the line width in lasers is connected with the phenomenon of spontaneous emission

Spatial directivity of the emission arises together with change in the spectral composition of the oscillation regime It is connected also with the nature of stimulated radiation during stimulated radiation, a light quantum has the same direction of propagation as the quantum which produced it

Usually in semiconductor lasers, the sample itself serves as a cavity, since semiconductor crystals have a large dielectric constant, and, since the polished boundary of the division between the air and the dielectric is capable of reflecting about 30% of the radiation

The first semiconductor lasers were made utilizing *p–n* junctions in crystals of GaAs [17, 18] Some time later, lasers were made under excitation by an electron beam [15], and, recently, under excitation by a light beam [23] In Table 1 different semiconducting materials are shown with which lasers have been made, and the methods of excitation are given

Table 1 Semiconductor Lasers

The semiconductor material	The wave range of the radiation (in microns)	The method of the excitation	References
CdS	0 5	high-speed electron beam	15
CdTe	0 8	high-speed electron beam	30
GaAs	0 85	*p–n* junction	17, 18
		high-speed electron beam	29
		optical excitation	23
InP	0 9	*p–n* junction	31
GaSb	1 6	*p–n* junction	32
		high-speed electron beam	33
InAs	3 2	*p–n* junction	34
		high-speed electron beam	35
InSb	5 3	*p–n* junction	36
		high-speed electron beam	6
PbTe	6 5	*p–n* junction	24
PbSe	8 5	*p–n* junction	24
GaAs–GaP	0 65–0 9	*p–n* junction	37
InAs–InP	0 9–3 2	*p–n* junction	25
GaAs–InAs	0 85–3 2	*p–n* junction	38

With the help of semiconductors it has already become possible to cover a large frequency range from 0.5 μ to 8.5 μ. In a number of cases it is possible to continuously overlap a very large frequency range, since the variation of the concentration of components in three-component semiconductors units results in changes in the distances between the bands, that is, allows one to continuously change the emission frequency. For instance, variation of composition in the system InAs–InP results in frequency changes from 0.9 μ to 3.2 μ [25].

At present, the highest degree of development has been obtained with lasers utilizing *p–n* junctions in GaAs. Impulse and continuous regimes were obtained with an average power of several watts, and peak power of up to 100 watts, with an efficiency of about 30% [24].

The most interesting characteristic of semiconductor lasers is their high efficiency.

Since a direct transformation of electric current into coherent emission takes place in lasers utilizing *p–n* junctions, their efficiency may approach unity. Even now it has become possible to make diodes with an efficiency of 70–80% [26].

Lasers with monochromatic optical pumping should also have a very high efficiency, since the pumping frequency may be made close to the emission frequency [11].

The efficiency of lasers with electron excitation cannot be higher than about 30% [12], since two-thirds of the energy is spent on heating of the lattice during the production of electron–hole pairs. However, such lasers may be rather powerful. This type of excitation will evidently make it possible to create sources of coherent emission working in the far ultraviolet range.

Another characteristic of semiconductors is a high coefficient of amplification, attaining a value of several thousands of reverse centimeters, which makes it possible to construct lasers with dimensions measured in microns, that is, with cavity dimensions close to the length of the emission wave. Such cavities should have a very short setup time, of the order of 10^{-12}–10^{-13} sec, which opens the way for the control of high frequencies by using the oscillations in semiconductor lasers, and for the creation of superfast-operating circuits on the basis of lasers, such as components for superfast-operating electronic computers. *Q*-switch lasers giving very short light pulses may be built out of semiconducting materials.

The small dimensions of semiconductor lasers make it possible to construct quantum amplifiers with an extremely high sensitivity, since sensitivity increases with a decrease in the number of modes of oscillation which may be excited in the cavity. For the first time light amplifiers with an amplification index of about 2000 cm^{-1} have been produced [28].

The high amplification index in semiconductor lasers makes it possible to create for them a new type of cavity—the cavity with emitting mirrors (Fig 8) [27]

A silver mirror is covered by a thin semiconductor film which is then covered by a transparent film If one produces in the semiconducting film a state with negative temperature which can compensate for the mirror losses, such a mirror may be used in the construction of a laser As in the case of a gas laser, one may expect to observe very high monochromaticity and spatial coherence in the emission A significant advantage of such a system is the relative ease of removing heat from the thin semiconducting film, which indicates that it should be possible to obtain considerable power

In order to produce negative temperature in a semiconducting film, one may use electronic excitation or optical pumping The utilization of semiconducting lasers with *p–n* junctions for optical pumping makes it possible to attain high efficiency in the system as a whole

The question as to the maximum power which may be obtained using semiconductor lasers is not quite clear at present However, the employment of emitting mirrors of sufficiently large area will make it possible, apparently, to utilize a considerable quantity of semiconducting material The maximum value of a mirror's cross-section is determined by such factors as the precision of its manufacture and the homogeneity of its semiconducting layer Various deviations from optical homogeneity will produce the highest modes of oscillations

Among the disadvantages of semiconductor lasers are their relatively small power, their large spatial divergence and their insufficiently high monochromaticity

However, in speaking about those disadvantages one should keep in mind that the field of semiconductor quantum electronics is still in its infancy Furthermore, the means of overcoming these disadvantages are already in sight It is quite clear in what directions to proceed in order to develop semiconductor quantum electronics, and to increase the sphere of application of semiconductor lasers All of this gives reason to hope that semiconductor quantum electronics will continue to play a fundamental role in the development of lasers

References

1 N G Basov, B M Vul, and Yu M Popov, *Sov Phys JETP* **37**, 585 (1959)
2 R H Dicke *Quantum Electronics*, p 572, Columbia University Press, New York (1960)
3 N G Basov, O N Krokhin, and Yu M Popov, *Usp Fiz Nauk* **72**, 161 (1960)
4 N G Basov, O N Krokhin, and Yu M Popov, *Sov Phys JETP* **39**, 1001 (1960)
5 W P Dumke, *Phys Rev* **127**, 1559 (1962)

6 C Benoit a la Guillaume and J M Debever, Symposium on Radiative Recombination in Semiconductors, Paris, July 28, 1964
7 N G Basov and A M Prochorov, *Sov Phys JETP* **28**, 249 (1955)
8 N Bloembergen, *Phys Rev* **104**, 324 (1956)
9 T H Maiman, *Nature* **187**, 493 (1960)
10 O N Krokhin and Yu M Popov, *Sov Phys JETP* **38**, 1589 (1960)
11 N G Basov and O N Krokhin, *Sov Phys JETP* **46**, 1508 (1964)
12 Yu M Popov *Proc FIAN* **XXIII**, 67 (1963)
13 V S Vavilov, *Usp Fiz Nauk* **XXV**, 263 (1961)
14 N G Basov, O N Krokhin, and Yu M Popov, *Adv Quantum Electron*, 496
15 N G Basov O V Bogdankevich and A Devyatkov, *Dokl Akad Nauk SSSR* **55**, 783 (1964)
16 N G Basov O N Krokhin, and Yu M Popov, *Sov Phys JETP* **40**, 1879 (1961)
17 R N Hall, G E Fenner, J D Kingsley, T J Soltys, and R O Carlson, *Phys Rev Lett* **9**, 366 (1962)
18 M I Nathan, W P Dumke, G Burns, F H Dill, and G J Lasher, *Appl Phys Lett* **1**, 62 (1962)
19 M Bernard (private communication)
20 A M Prokhorov, *Sov Phys JETP* **34**, 1658 (1959)
21 A L Schawlow and C H Townes, *Phys Rev* **112**, 1940 (1958)
22 J A Armstrong and A W Smith, *Appl Phys Lett* **4**, 196 (1964)
23 N G Basov, A Z Grasjuk and V A Katulin, *Dokl Akad Nauk SSSR* **161**, 1306 (1965)
24 C Hilsum, "Lasers and their application, London, September 1964
25 F B Alexander, *Appl Phys Lett* **4**, 13 (1964)
26 M I Nathan, *Proc IEEE* **52**, 770 (1964)
27 M I Basov and O V Bogdankevich, Symposium on Radiative Recombination in Semiconductors Paris, July 28, 1964
28 J W Crowe and R W Craig, *Appl Phys Lett* **4**, 57 (1964)
29 C E Hurwitz and R J Keyes, *Appl Phys Lett* **5**, 139 (1964)
30 V S Vavilov, E L Nolle, and V D Egorov, *Fiz Tverd Tela* **7**, 934 (1965)
31 G Burns, R S Levitt, M I Nathan and K Weiser *Proc EEE* **51**, 1148 (1963)
32 T Deutsch *et al*, *Phys Status Solidi* **3**, 1001 (1963)
33 G C Benoit a la Guillaume and J M Debever *Compt Rend* **259**, 2200 (1964)
34 I Melngilis *Appl Phys Lett* **2**, 176 (1963)
35 G C Benoit a la Guillaume and J M Debever *Solid State Commun* **2**, 145 (1964)
36 R J Phelan, A R Calawa, R H Rediker, R J Keyes, and B Lax, *Appl Phys Lett* **3**, 143 (1963)
37 N Holonyak Jr and S F Bevacqua, *Appl Phys Lett* **1**, 82 (1962)
38 T M Quist, R H Rediker R J Keyes and W E Prag *Bull Am Phys Soc* **January** 88 (1963)

Nobel Lecture, December 11, 1971

Holography, 1948–1971

Dennis Gabor

I have the advantage in this lecture, over many of my predecessors, that I need not write down a single equation or show an abstract graph One can of course introduce almost any amount of mathematics into holography, but the essentials can be explained and understood from physical arguments

Holography is based on the wave nature of light, and this was demonstrated convincingly for the first time in 1801 by Thomas Young, by a wonderfully simple experiment (Figure 1) He let a ray of sunlight into a dark room, placed a dark screen in front of it, pierced with two small pinholes, and beyond this, at some distance, a white screen He then saw two darkish lines at both sides of a bright line, which gave him sufficient encouragement to repeat the experiment, this time with a spirit flame as light source, with a little salt in it, to produce the bright yellow sodium light This time he saw a number of dark lines, regularly spaced, the first clear proof that light added to light can produce darkness This phenomenon is called interference Thomas Young had expected it because he believed in the wave theory of light His great contribution to Christian Huygens's original idea was the intuition that monochromatic light represents regular, sinusoidal oscillations, in a medium which at that time was called "the ether " If this is so, it must be possible to produce more light by adding wavecrest to wavecrest, and darkness by adding wavecrest to wavetrough

Light which is capable of interference is called "coherent," and it is evident that in order to yield many interference fringes, it must be *very* monochromatic Coherence is conveniently measured by the path difference between two rays of the same source, by which they can differ while

Dennis Gabor • Imperial Colleges of Science and Technology, London

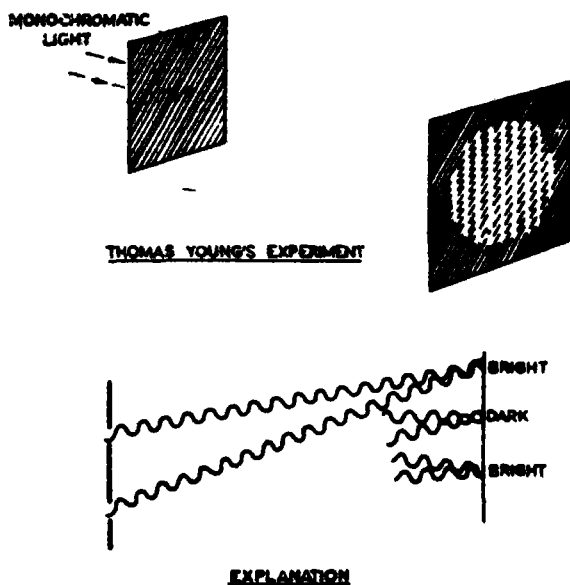



Fig 1 Thomas Young's interference experiments 1801

still giving observable interference contrast This is called the coherence length, an important quantity in the theory and practice of holography Lord Rayleigh and Albert Michelson were the first to understand that it is a reciprocal measure of the spectroscopic line width Michelson used it for ingenious methods of spectral analysis and for the measurement of the diameter of stars

Let us now jump a century and a half, to 1947 At that time I was very interested in electron microscopy This wonderful instrument had at that time produced a hundredfold improvement on the resolving power of the best light microscopes, and yet it was disappointing, because it had stopped short of resolving atomic lattices The de Broglie wavelength of fast electrons, about 1/20 Å, was short enough, but the optics was imperfect The best electron objective which one can make can be compared in optical perfection to a raindrop than to a microscope objective, and through the theoretical work of O Scherzer it was known that it could never be perfected The theoretical limit at that time was estimated at 4 Å, just about twice what was needed to resolve atomic lattices, while the practical limit stood at about 12 Å These limits were given by the necessity of restricting the aperture of the electron lenses to about 1/1000 radian, at which angle the spherical aberration error is

about equal to the diffraction error If one doubles this aperture so that the diffraction error is halved, the spherical aberration error is increased 8 times, and the image is hopelessly blurred

After pondering this problem for a long time, a solution suddenly dawned on me, one fine day at Easter 1947, more or less as shown in Figure 2 Why not take a bad electron picture, but one which contains the *whole* information, and correct it by optical means? It was clear to me for some time that this could be done, if at all, only with coherent electron beams, with electron waves which have a definite phase But an ordinary photograph loses the phase completely, it records only the intensities No wonder we lose the phase, if there is nothing to compare it with! Let us see what happens if we add a standard to it, a "coherent background" My argument is illustrated in Figure 2, for the simple case when there is only one object point The interference of the object wave and of the coherent background or "reference wave" will then produce interference fringes There will be maxima wherever the phases of the two waves were identical Let us make a hard positive record, so that it transmits only at the maxima, and illuminate it with the reference source alone Now the

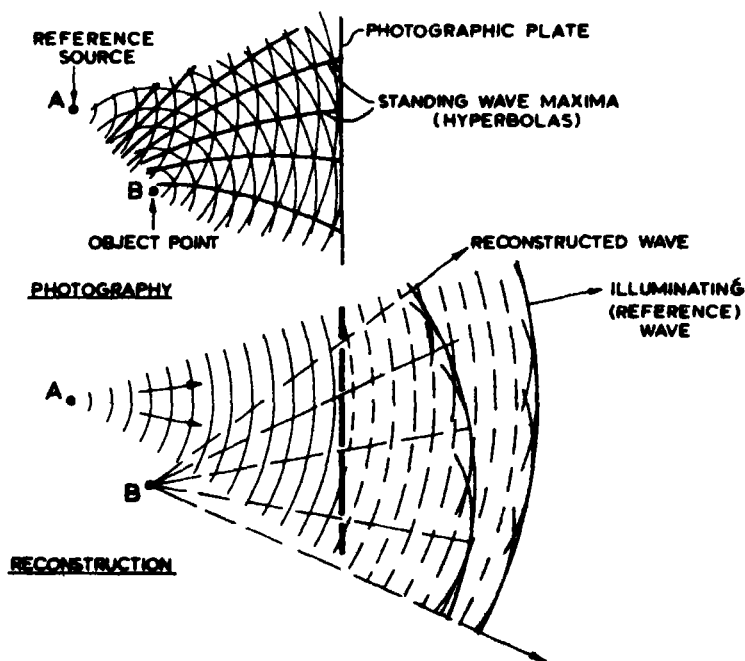



Fig 2 The basic idea of holography, 1947

phases are of course right for the reference source *A*, but as at the slits the phases are identical, they must be right also for *B*, therefore the wave of *B* must also appear, *reconstructed*

A little mathematics soon showed that the principle was right, also for more than one object point, for any complicated object Later on it turned out that in holography Nature is on the inventor's side, there is no need to take a hard positive record, one can take almost any negative This encouraged me to complete my scheme of electron microscopy by reconstructed wavefronts, as I then called it and to propose the two-stage process shown in Figure 3 The electron microscope was to produce the interference figure between the object beam and the coherent background, that is to say the non-diffracted part of the illuminating beam This interference pattern I called a "hologram," from the Greek word "holos"—the whole, because it contained the whole information The hologram was then reconstructed with light, in an optical system which corrected the aberrations of the electron optics [1]

In doing this, I stood on the shoulders of two great physicists, W L Bragg and Frits Zernike Bragg had shown me, a few years earlier, his "X-ray microscope" an optical Fourier-transformer device One puts into it a small photograph of the reciprocal lattice, and obtains a projection of the electron densities, but only in certain exceptional cases, when the phases are all real, and have the same sign I did not know at that time, and neither did Bragg, that Mieczislav Wolfke had proposed this method

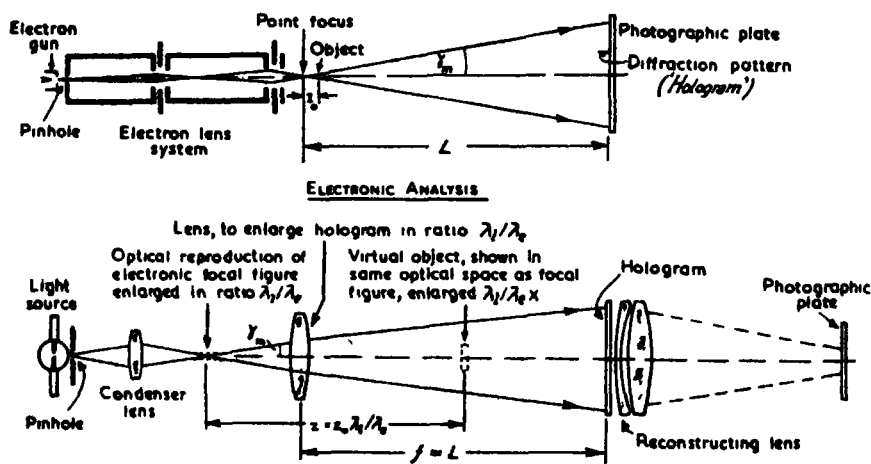



Fig 3 The principle of electron microscopy by reconstructed wavefronts (Gabor, *Proc R Soc London Ser A* **197**, 454, 1949)

in 1921, but without realising it experimentally.† So the idea of a two-stage method was inspired by Bragg. The coherent background, on the other hand, was used with great success by Frits Zernike in his beautiful investigations of lens aberrations, showing up their phase, and not just their intensity. It was only the reconstruction principle which had escaped them.

In 1947 I was working in the Research Laboratory of the British Thomson–Houston Company in Rugby, England. It was a lucky thing that the idea of holography came to me *via* electron microscopy, because if I had thought of optical holography only, the Director of Research, L. J. Davies, could have objected that the BTH company was an electrical engineering firm, and not in the optical field. But as our sister company, Metropolitan Vickers were makers of electron microscopes, I obtained the permission to carry out some optical experiments. Figure 4 shows one of our first holographic reconstructions. The experiments were not easy. The best compromise between coherence and intensity was offered by the high pressure mercury lamp, which had a coherence length of only 0.1 mm, enough for about 200 fringes. But in order to achieve spatial coherence, we (my assistant Ivor Williams and I) had to illuminate, with one mercury line, a pinhole of 3 microns diameter. This left us with enough light to make holograms of about 1 cm diameter of objects, which were microphotographs of about 1 mm diameter, with exposures of a few minutes, on the most sensitive emulsions then available. The small coherence length forced us to arrange everything in one axis. This is now called "in-line" holography, and it was the only one possible at that time. Figure 5 shows a somewhat improved experiment, the best of our series. It was far from perfect. Apart from the *schlieren*, which cause random disturbances, there was a systematic defect in the pictures, as may be seen by the distortion of the letters. The explanation is given in Figure 6. The disturbance arises from the fact that there is not one image but *two*. Each point of the object emits a spherical secondary wave, which interferes with the background and produces a system of circular Fresnel zones. Such a system is known after the optician who first produced it, a Soret lens. This is, at the same time, a positive and a negative lens. One of its foci is in the original position of the object point, the other in a position conjugate to it, with respect to the illuminating wavefront. If one uses "in-line holography" both images are in line, and can be separated only by focusing. But the separation is never quite perfect, because in regular, coherent illumination every point leaves a "wake" behind it, which reaches to long distances.

† M. Wolfke, *Phys. Z.* **21**, 495–497 (*Sept.* 15, 1920)

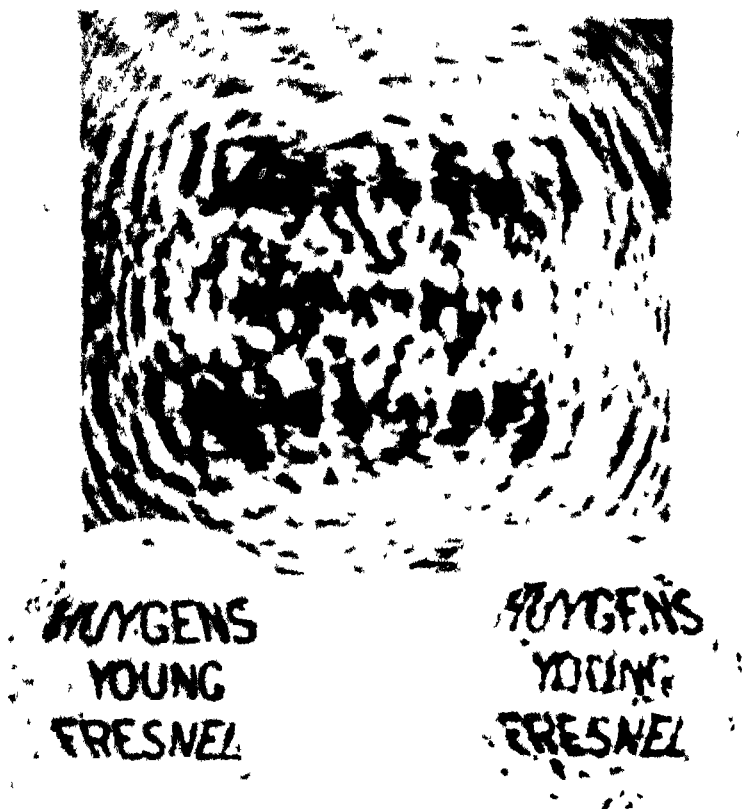



Fig 4 First holographic reconstruction, 1948

I will tell later with what case modern laser holography has got rid of this disturbance, by making use of the superior coherence of laser light which was not at my disposal in 1948 However, I was confident that I could eliminate the second image in the application which alone interested me at that time seeing atoms with the electron microscope This method, illustrated in Figure 7, utilized the very defect of electron lenses, the spherical aberration, in order to defeat the second image If an electron hologram is taken with a lens with spherical aberration, one can afterwards correct *one* of the two images by suitable optics, and the other has then twice the aberration, which washes it out almost completely

Figure 7 shows that a perfectly sharp reconstruction, in which as good as nothing remains of the disturbance caused by the second image, can be obtained with a lens so bad that its definition is at least 10 times worse than the resolution which one wants to obtain Such a very bad lens was obtained using a microscope objective the wrong way round, and using it again in the reconstruction

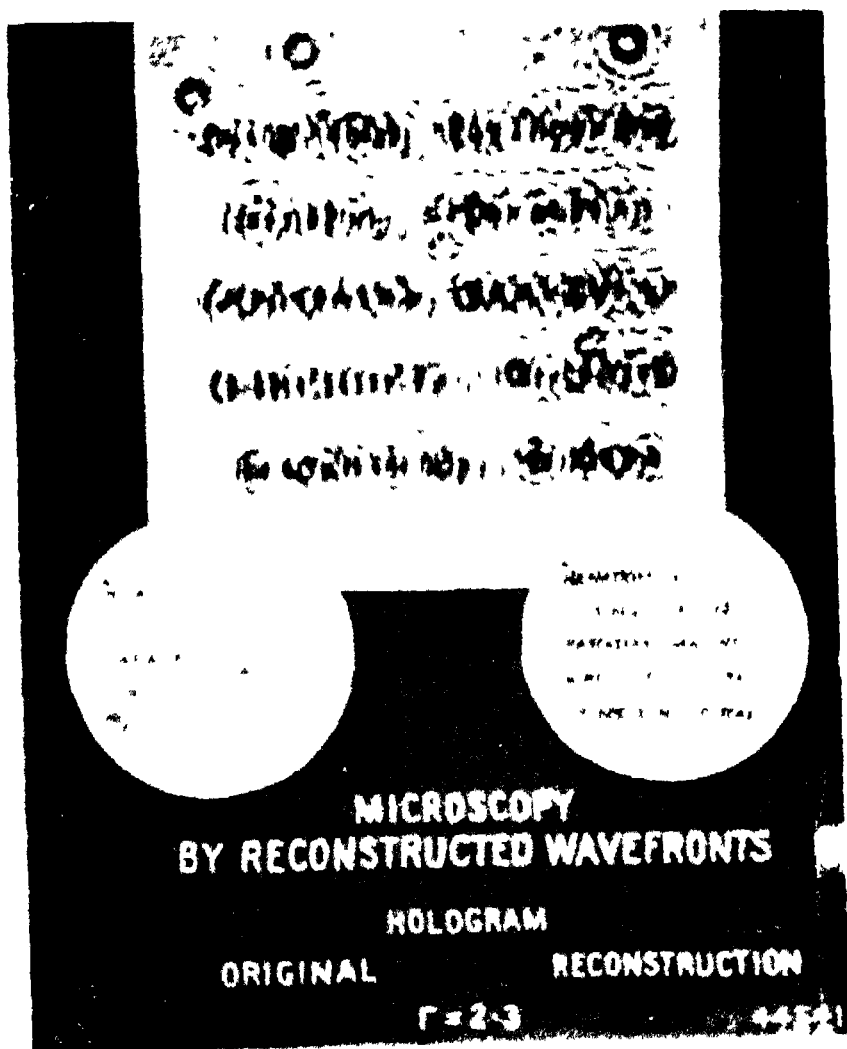



Fig ~ Another example of early holography 1948 (Gabor *Proc R Soc London Ser A* 197 454 1949)

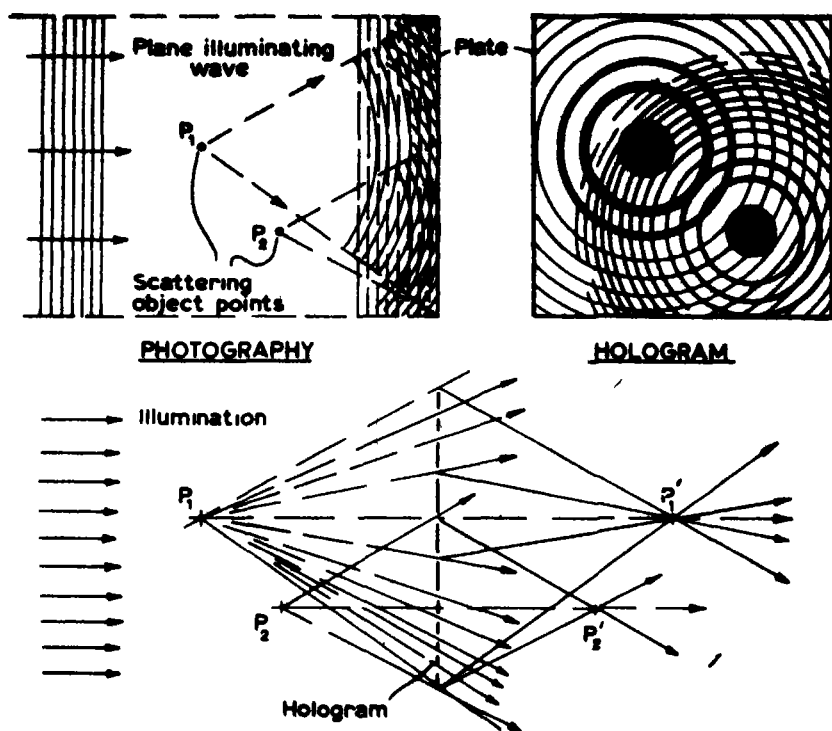



Fig 6 The second image Explanation in terms of Soret-lenses as holograms of single object points

So it was with some confidence that two years later, in 1950 we started a programme of holographic electron microscopy in the Research Laboratory of the Associated Electrical Industries, in Aldermaston, under the direction of Dr T E Allibone, with my friends and collaborators M W Haine, J Dyson and T Mulvey † By that time I had joined Imperial College, and took part in the work as a consultant In the course of three years we succeeded in considerably improving the electron microscope, but in the end we had to give up, because we had started too early It turned out that the electron microscope was still far from the limit imposed by optical aberrations It suffered from vibrations, stray magnetic fields, creep of the stage, contamination of the object, all made worse by the long exposures required in the weak coherent electron beam Now, 20

† Supported by a grant of the D S I R (Direction of Scientific and Industrial Research) the first research grant ever given by that body to an industrial laboratory

years later, would be the right time to start on such a programme, because in the meantime the patient work of electron microscopists have overcome all these defects The electron microscope resolution is now right up to the limit set by the spherical aberration, about 3 5 Å, and only an improvement by a factor of 2 is needed to resolve atomic lattices Moreover, there is no need now for such very long exposures as we had to contemplate in 1951, because by the development of the field emission cathode the coherent current has increased by a factor of 3–4 orders of magnitude So perhaps I may yet live to see the realisation of my old ideas

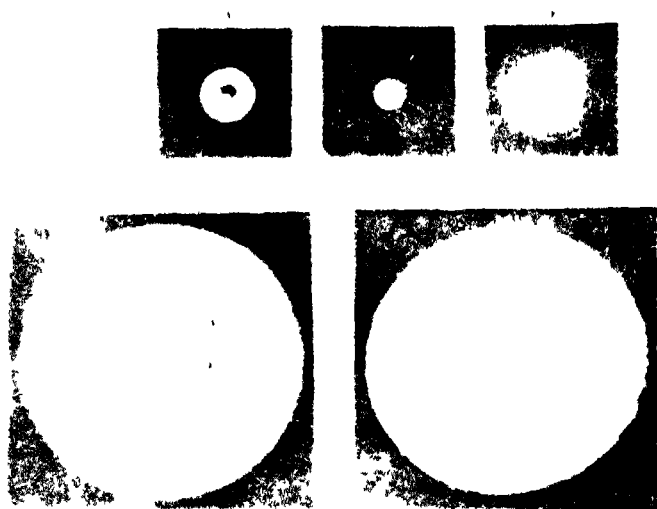



Fig 7 Elimination of the second image by compensation of the spherical aberration in the reconstruction [Gabor 1948 published 1951 (1)]

My first papers on wavefront reconstruction evoked some immediate responses G L Rogers [2] in Britain made important contributions to the technique, by producing among other things the first phase holograms, and also by elucidating the theory In California Alberto Baez [3], Hussein El-Sum and P Kirkpatrick [4] made interesting forays into X-ray holography For my part, with my collaborator W P Goss, I constructed a holographic interference microscope, in which the second image was annulled in a rather complicated way by the superimposition of two holograms, "in quadrature" with one another The response of the optical industry to this was so disappointing that we did not publish a paper on it until 11 years later, in 1966 [5] Around 1955 holography went into a long hybernation

The revival came suddenly and explosively in 1963, with the publication of the first successful laser† holograms by Emmett N Leith and Juris Upatnieks of the University of Michigan, Ann Arbor Their success was due not only to the laser, but to the long theoretical preparation of Emmett Leith, which started in 1955 This was unknown to me and to the world, because Leith, with his collaborators Cutrona, Palermo, Porcello and Vivian applied his ideas first to the problem of the "side-looking radar" which at that time was classified [6] This was in fact two-dimensional holography with electromagnetic waves, a counterpart of electron holography The electromagnetic waves used in radar are about 100,000 times longer than light waves, while electron waves are about 100,000 times shorter Their results were brilliant, but to my regret I cannot discuss them for lack of time

When the laser became available, in 1962, Leith and Upatnieks could at once produce results far superior to mine, by a new, simple and very effective method of eliminating the second image [7] This is the method of the "skew reference wave," illustrated in Fig 8 It was made possible by the great coherence length of the helium–neon laser, which even in 1962 exceeded that of the mercury lamp by a factor of about 3000 This made it possible to separate the reference beam from the illuminating beam, instead of going through the object, it could now go around it The result was that the two reconstructed images were now

I have been asked more than once why I did not invent the laser In fact I have thought of it In 1950, thinking of the desirability of a strong source of coherent light, I remembered that in 1921 as a young student, in Berlin, I had heard from Einstein's own lips his wonderful derivation of Planck's law which postulated the existence of stimulated emission I then had the idea of the pulsed laser Take a suitable crystal make a resonator of it by a highly reflecting coating, fill up the upper level by illuminating it through a small hole and discharge it explosively by a ray of its own light I offered the idea as a Ph D problem to my best student but he declined it, as too risky, and I could not gainsay it as I could not be sure that we would find a suitable crystal

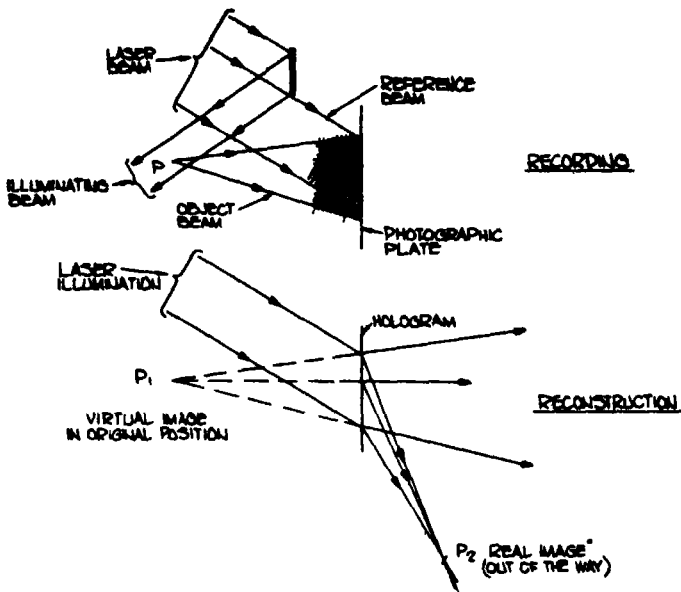



Fig 8 Holography with skew reference beam (E N Leith and J Upatnieks, 1963)

separated not only in depth, but also angularly, by twice the incidence angle of the reference beam Moreover, the intensity of the coherent laser light exceeded that of mercury many millionfold This made it possible to use very fine-grain, low speed photographic emulsions and to produce large holograms, with reasonable exposure times

Figure 9 shows two of the earliest reconstructions made by Leith and Upatnieks, in 1963, which were already greatly superior to anything that I could produce in 1948 The special interest of these two images is, that they are reconstructions from *one* hologram, taken with different positions of the reference beam This was the first proof of the superior storage capacity of holograms Leith and Upatnieks could soon store 12 different pictures in one emulsion Nowadays one can store 100 or even 300 pages of printed matter in an area which by ordinary photography would be sufficient for one

From then on progress became very rapid The most spectacular result of the first year was the holography of three-dimensional objects, which could be seen with two eyes Holography was of course three-dimensional from the start, but in my early, small holograms one could see this only by focusing through the field with a microscope or short-focus eyepiece But it was not enough to make the hologram large, it was

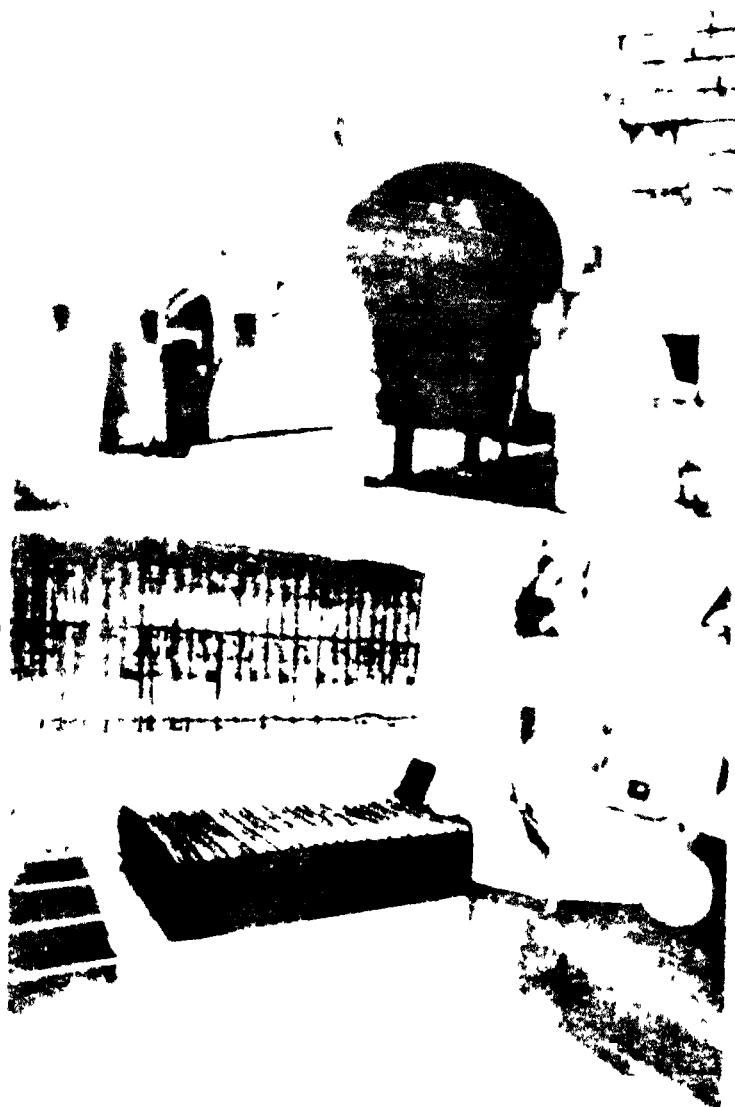



Fig. 9. First example of multiple image storage in one hologram (E. N. Leith and J. Upatnieks, *J. Opt. Soc. Am.* November 1964)

also necessary that every point of the photographic plate should see every point of the object In the early holograms, taken with regular illumination, the information was contained in a small area, in the diffraction pattern

In the case of rough, diffusing objects no special precautions are necessary The small dimples and projections of the surface diffuse the light over a large cone Figure 10 shows an example of the setup in the case of a rough object, such as a statuette of Abraham Lincoln The reconstruction is shown in Figure 11 With a bleached hologram ("phase hologram") one has the impression of looking through a clear window at the statuette itself

If the object is non-diffusing, for instance if it is a transparency, the information is spread over the whole hologram area by illuminating the object through a diffuser, such as a frosted glass plate The appearance of such a "diffused" hologram is extraordinary, it looks like noise (Fig 12)

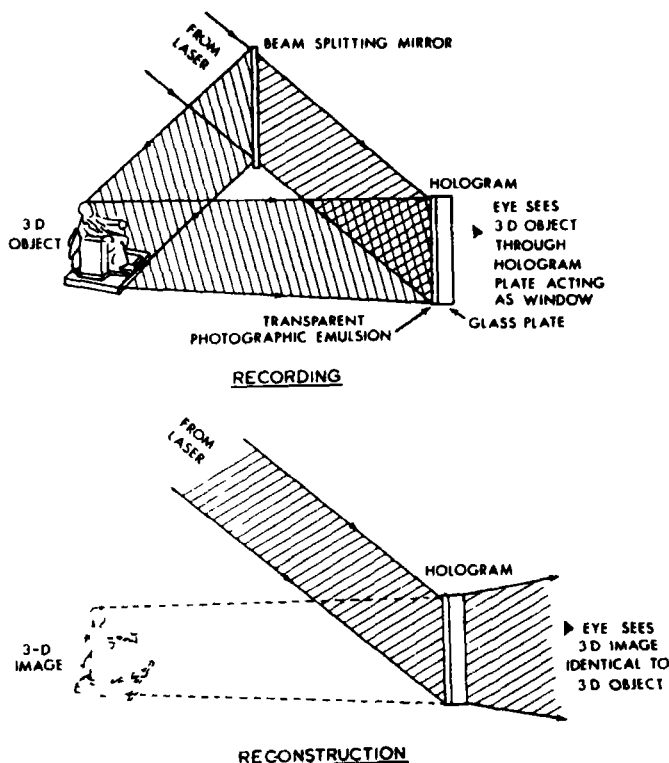



Fig 10 3-D Holography of a diffusing object with laser light

Fig 11 Three-dimensional reconstruction of a small statue of Abraham Lincoln (Courtesy of Professor G W Stroke, State University of New York, Stony Brook)

One can call it "ideal Shannon coding," because Claude E Shannon has shown in his Communication Theory that the most efficient coding is such that all regularities seem to have disappeared in the signal, it must be "noise-like " But where is the information in this chaos? It can be shown that it is not as irregular as it appears It is not as if grains of sand had been scattered over the plate at random It is rather a complicated figure, the diffraction pattern of the object, which is repeated at random intervals, but always in the same size and same orientation

A very interesting and important property of such diffused holograms is that any small part of it, large enough to contain the diffraction pattern, contains information on the whole object, and this can be reconstructed from the fragment, only with more noise A diffuse hologram is therefore a *distributed memory*, and this has evoked much speculation whether human memory is not perhaps, as it were, holographic, because it is well known that a good part of the brain can be destroyed without wiping out every trace of a memory There is no time here to discuss this very exciting question I want only to say that in my opinion the similarity with the human memory is functional only, but certainly not structural

It is seen that in the development of holography the hologram has become always more unlike the object, but the reconstruction always more perfect Figure 13 shows an excellent reconstruction by Leith and Upatnieks of a photograph, from a diffuse hologram like the one in the previous figure

The pioneer work carried out in the University of Michigan, Ann Arbor, led also to the stabilization of holographic techniques Today hundreds if not thousands of laboratories possess the equipment of which an example is shown in Figure 14, the very stable granite slab or steel table, and the various optical devices for dealing with coherent light, which are now manufactured by the optical industry The great stability is absolutely essential in all work carried out with steady-state lasers, because a movement of the order of a quarter wavelength during the exposure can completely spoil a hologram

Fig 12 Strongly magnified image of a hologram taken with diffused illumination The information is conveyed in a noiselike code (E N Leith and J Upatnieks, 1964)

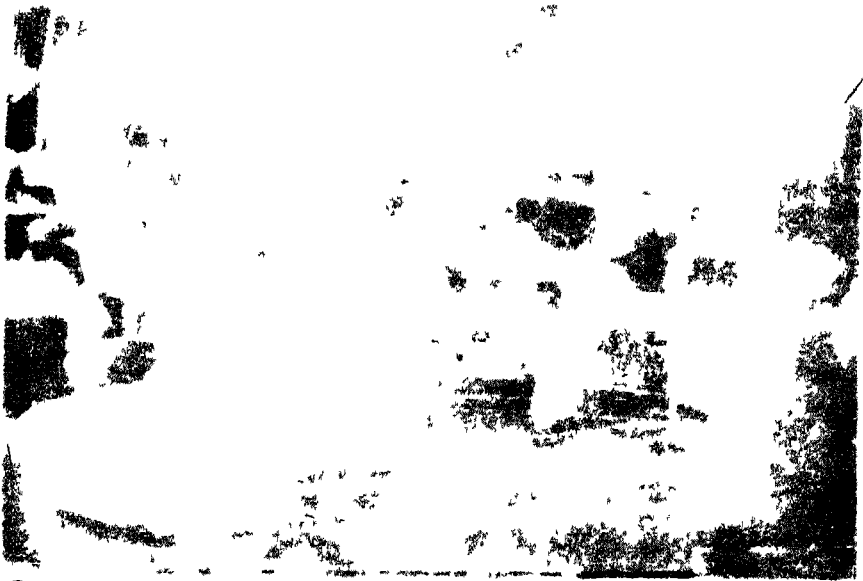



Fig 13 Reconstruction of a plane transparency showing a restaurant, from a hologram taken with diffused illuminations (E N Leith and J Upatnieks 1964)

However, from 1965 onwards there has developed an important branch of holography where high stability is not required, because the holograms are taken in a small fraction of a microsecond, with a pulsed laser

Imagine that you had given a physicist the problem "Determine the size of the droplets which issue from a jet nozzle, with a velocity of 2 Mach The sizes are probably from a few microns upwards " Certainly he would have thrown up his hands in despair! But all it takes now, is to record a simple in-line hologram of the jet, with the plate at a safe distance, with a ruby laser pulse of 20–30 nanoseconds One then looks at the "real" image (or one reverses the illuminating beam and makes a real image of the virtual one), one dives with a microscope into the three-dimensional image of the jet and focuses the particles, one after the other Because of the large distance, the disturbance by the second image is entirely negligible Figure 15 shows a fine example

As the research workers of the TRW laboratories have shown, it is possible to record in one hologram the infusoriae in several feet of dirty water, or insects in a meter of air space Figure 16 shows two reconstructions of insects from one hologram, focusing on one after the other The

authors, C Knox and R E Brooks, have also made a cinematographic record of a holographic film, in which the flight of one mosquito is followed through a considerable depth, by refocusing in every frame [9]

Another achievement of the TRW group, Ralph Wuerker and his colleagues, leads us into another branch of holography, to holographic interferometry Figure 17 shows a reconstruction of a bullet, with its train of shockwaves, as it meets another shockwave But is is not just an image, it is an interferometric image The fringes show the *loci* at which the retardation of light is by integer wavelengths, relative to the quiet air, before the event This comparison standard is obtained by a previous exposure This is therefore a double-exposure hologram, such as will be discussed in more detail later [10]

Figure 18 shows another high achievement of pulse holography a holographic, three-dimensional portrait, obtained by L Siebert in the Conductron Corporation (now merged into McDonnel-Douglas Electronics Company, St Charles, Missouri) It is the result of outstanding work in the development of lasers The ruby laser, as first realised by T H Maiman was capable of short pulses, but its coherence length was of

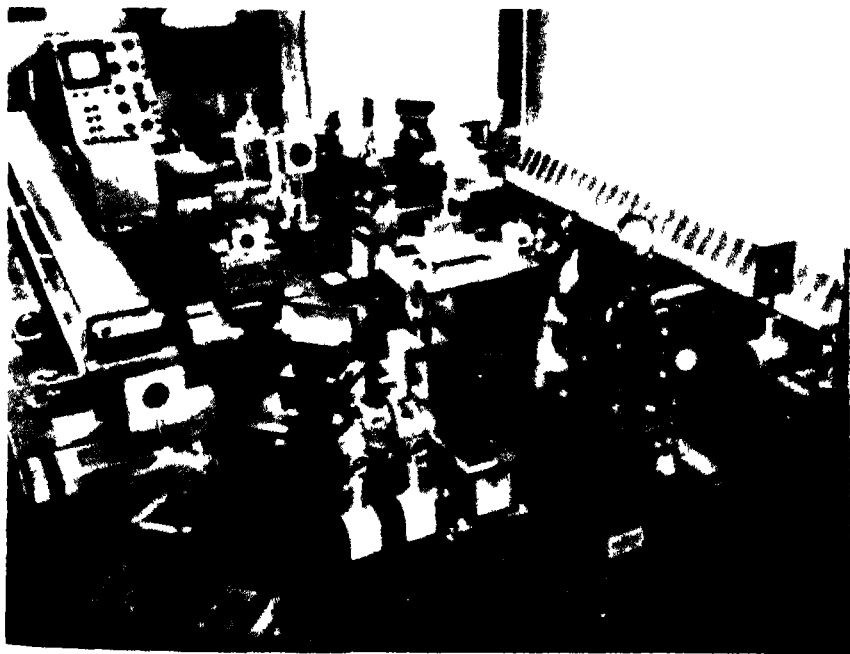



Fig 14 Modern holographic equipment

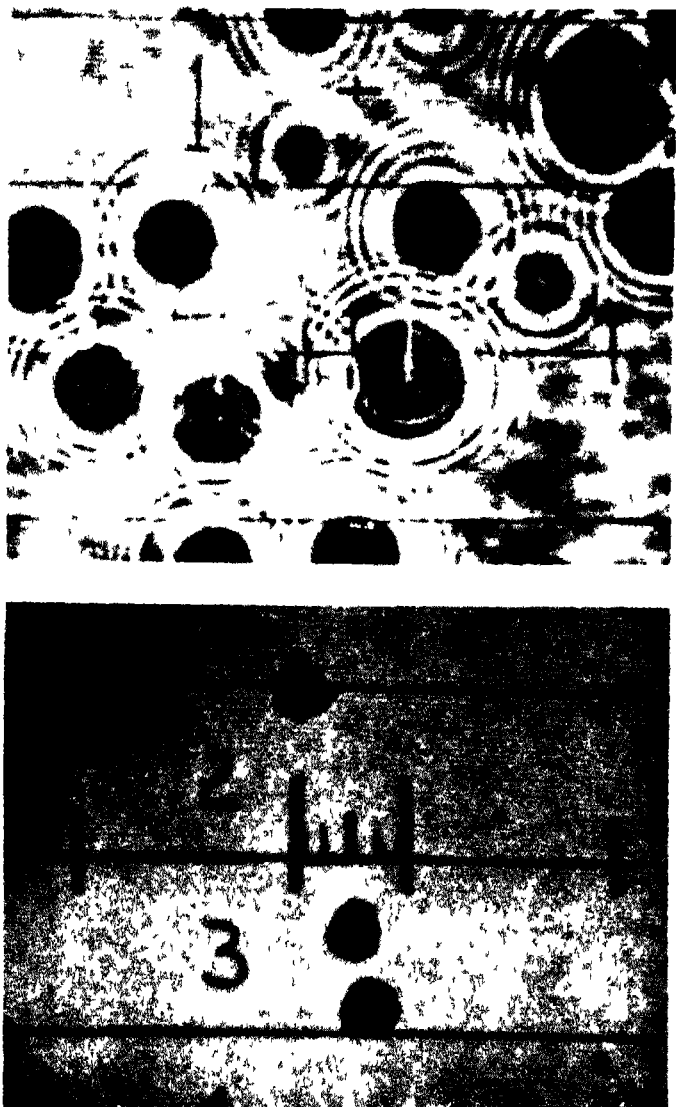



Fig 15 Holography of jets (Courtesy of Laser Holography Inc, Santa Barbara California)

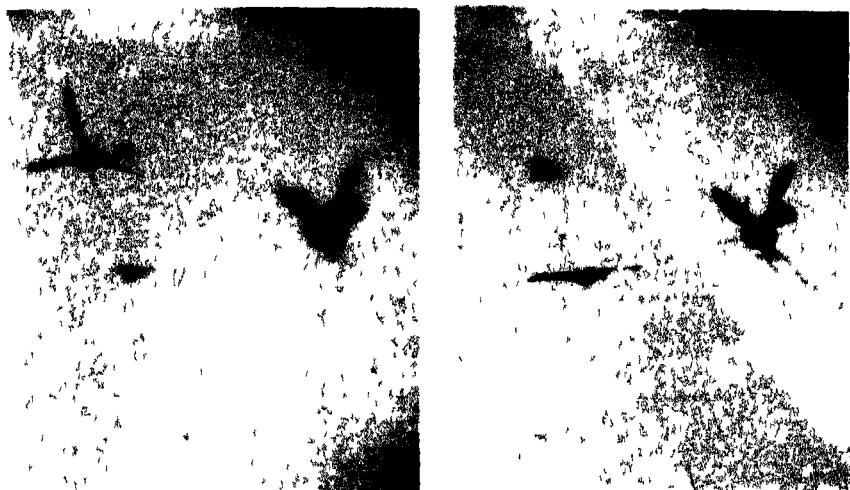



Fig. 16 Observation of mosquitos in flight Both pictures are extracted from one hologram (Courtesy of C Know and R E Brooks, TRW, Redondo Beach, California.)

the order of a few cm only This is no obstacle in the case of in-line holography, where the reference wave proceeds almost in step with the diffracted wavelets, but in order to take a scene of, say, one meter depth with reflecting objects one must have a coherence length of at least one meter Nowadays single-mode pulses of 30 nanosecond duration with 10 joule in the beam and coherence lengths of 5–8 meters are available, and have been used recently for taking my holographic portrait shown in the exhibition attached to this lecture [the portrait appears on page 400]

In 1965 R L Powell and K A Stetson at the University of Michigan, Ann Arbor, made an interesting discovery Holographic images taken of moving objects are washed out But if double exposure is used, first with the object at rest, then in vibration, fringes will appear, indicating the lines where the displacement amounted to multiples of a half wavelength Figure 19 shows vibrational modes of a loudspeaker membrane, recorded in 1965 by Powell and Stetson [11], Fig 20 the same for a guitar, taken by K A Stetson in the laboratory of Professor Erik Ingelstam [12]

Curiously, both the interferograms of the TRW group and the vibrational records of Powell and Stetson preceded what is really a simpler application of the interferometrical principle, and which historically ought to have come first—if the course of science would always follow the shortest line This is the observation of small deformations of

Fig 17 Dynamic holographic interferometry This reconstruction of a holographic interferogram shows the interaction of two air shock fronts and their associated flows (Courtesy of Dr R F Wuerker and his associates, TRW Physical Electronics Laboratory, Redondo Beach, California)

solid bodies, by double exposure holograms A simple explanation is as follows We take a hologram of a body in State *A* This means that we freeze in the wave *A* by means of a reference beam Now let us deform the body so that is assumes the State *B* and take a second hologram in the same emulsion with the same reference beam We develop the hologram, and illuminate it with the reference beam Now the two waves

Fig 18 Holographic portrait (L Siebert, Conductron Corporation, now merged into McDonnell-Douglas Electronics Company, St Charles, Missouri)

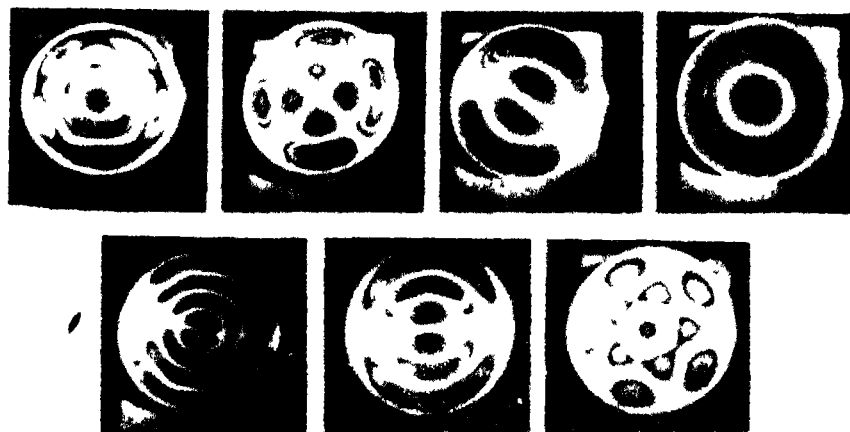



Fig 19 Vibrational modes of a loudspeaker membrane, obtained by holographic interferometry (R L Powell and K A Stetson, University of Michigan Ann Arbor 1965)

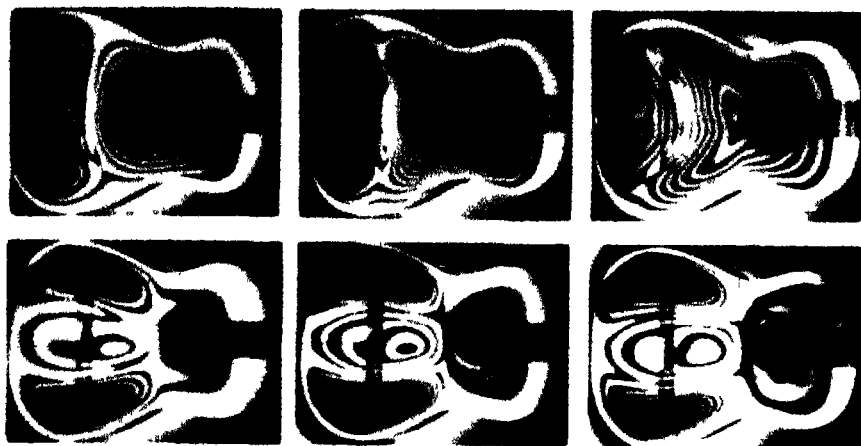



Fig 20 Vibrational modes of a guitar, recorded by holographic interferometry (Courtesy of Dr K A Stetson and Professor E Ingelstam)

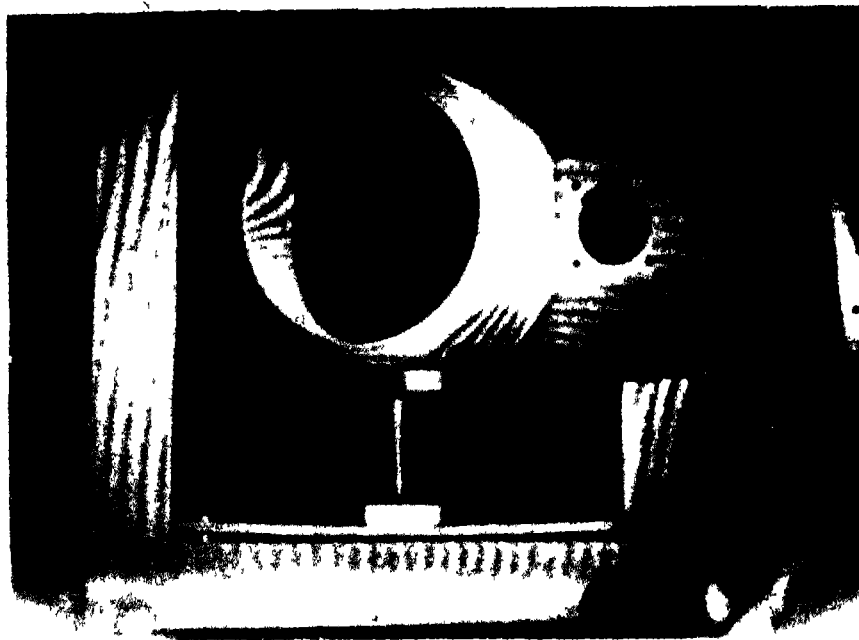



Fig 21 An early example of holographic interferometry by double exposure (Haines and Hildebrand University of Michigan Ann Arbor 1965)

A and *B*, frozen in at different times, and which have never seen one another, will be revived simultaneously, and they interfere with one another. The result is that Newton fringes will appear on the object, each fringe corresponding to a deformation of a half wavelength Figure 21 shows a fine example of such a holographic interferogram, made in 1965 by Haines and Hildebrand The principle was discovered simultaneously and independently also by J M Burch in England, and by G W Stroke and A Labeyrie in Ann Arbor, Michigan

Non-destructive testing by holographic interferometry is now by far the most important industrial application of holography It gave rise to the first industrial firm based on holography, GCO (formerly G C Optronics), in Ann Arbor, Michigan, and the following examples are reproduced by courtesy of GCO Figure 22 shows the testing of a motor car tyre The front of the tyre is holographed directly, the sides are seen in two mirrors, right and left First a little time is needed for the tyre to

Fig 22 Non-destructive testing by holography Double exposure hologram revealing two flaws in a tyre (Courtesy of Dr Ralph Grant and GCO Ann Arbor Michigan)

settle down and a first hologram is taken Then a little hot air is blown against it, and a second exposure is made, on the same plate If the tyre is perfect, only a few, widely spaced fringes will appear, indicating almost uniform expansion But where the cementing of the rubber sheets was imperfect, a little blister appears, as seen near the centre and near the top left corner, only a few thousandths of a millimeter high, but indicating a defect which could become serious Alternatively, the first hologram is developed, replaced exactly in the original position, and the expansion of the tyre is observed "live"

Other examples of non-destructive testing are shown in Fig 23, all defects which are impossible or almost impossible to detect by other means, but which reveal themselves unmistakeably to the eye A particularly impressive piece of equipment manufactured by GCO is shown in Fig 24 It is a holographic analyser for honeycomb sandwich structures (such as shown in the middle of Fig 23) which are used in aeroplane wings The smallest welding defect between the aluminum sheets and the honeycomb is safely detected at one glance

While holographic interferometry is perfectly suited for the detection

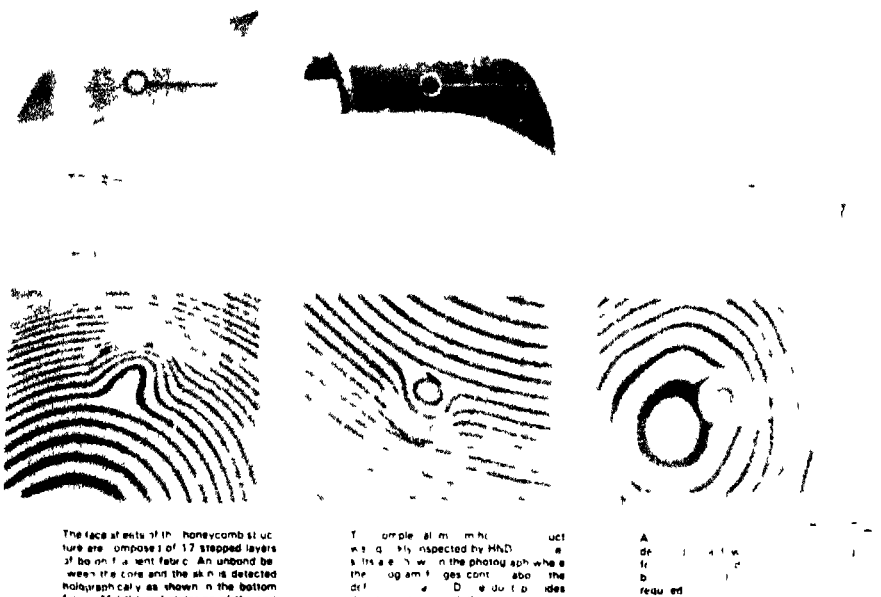



Fig 23 Examples of holographic non-destructive testing (Courtesy of GCO Ann Arbor Michigan)

Fig. 24 Holographic analyzer Mark II for sandwich structures GCO Ann Arbor Michigan

of very small deformations, with its fringe unit of 1/4000 mm, it is a little too fine for the checking of the accuracy of workpieces Here another holographic technique called "contouring" is appropriate It was first introduced by Haines and Hildebrand, in 1965, and has been recently much improved by J Varner, also in Ann Arbor, Michigan Two holograms are taken of the same object, but with two wavelengths which differ by e g one percent This produces *beats* between the two-fringe system, with fringe spacings corresponding to about 1/40 mm, which is just what the workshop requires (Fig 25)

From industrial applications I am now turning to another important development in holography In 1962, just before the "holography explosion" the Soviet physicist Yu N Denisyuk published an important paper [13] in which he combined holography with the ingenious method of photography in natural colours, for which Gabriel Lippman received the Nobel Prize in 1908 Figure 26a illustrates Lippmann's method and Denisyuk's idea Lippmann produced a very fine-grain emulsion, with colloidal silver bromide, and backed the emulsion with mercury, serving as a mirror Light falling on the emulsion was reflected at the mirror, and produced a set of standing waves Colloidal silver grains were precipitated

Fig 25 Holographically produced contour map of a medal made by a meth initiated by B P Hildebrand and K Haines (*J Opt Soc Am* 57, 155 1967) Improved by J Varne, University of Michigan Ann Arbor 1969

in the maxima of the electric vector, in layers spaced by very nearly half a wavelength After development, the complex of layers, illuminated with white light, reflected only a narrow waveband around the original colour, because only for this colour did the wavelets scattered at the Lippmann layers add up in phase

Denisyuk's suggestion is shown in the second diagram (Fig 26b) The object wave and the reference wave fall in from opposite sides of the emulsion Again standing waves are produced, and Lippman layers, but these are no longer parallel to the emulsion surface, they bisect the angle between the two wavefronts If now, and this is Denisyuk's principle, the developed emulsion is illuminated by the reference wave, the object will appear, in the original position and (unless the emulsion has shrunk) in the original colour

Though Denisyuk showed considerable experimental skill, lacking a laser in 1962 he could produce only an "existence proof " A two-colour

reflecting hologram which could be illuminated with white light was first produced in 1965 by G W Stroke and A Labeyrie [14] and is shown in Fig 27

Since that time single-colour reflecting holograms have been developed to high perfection by new photographic processes, by K S Pennington [15] and others, with reflectances approaching 100 percent,

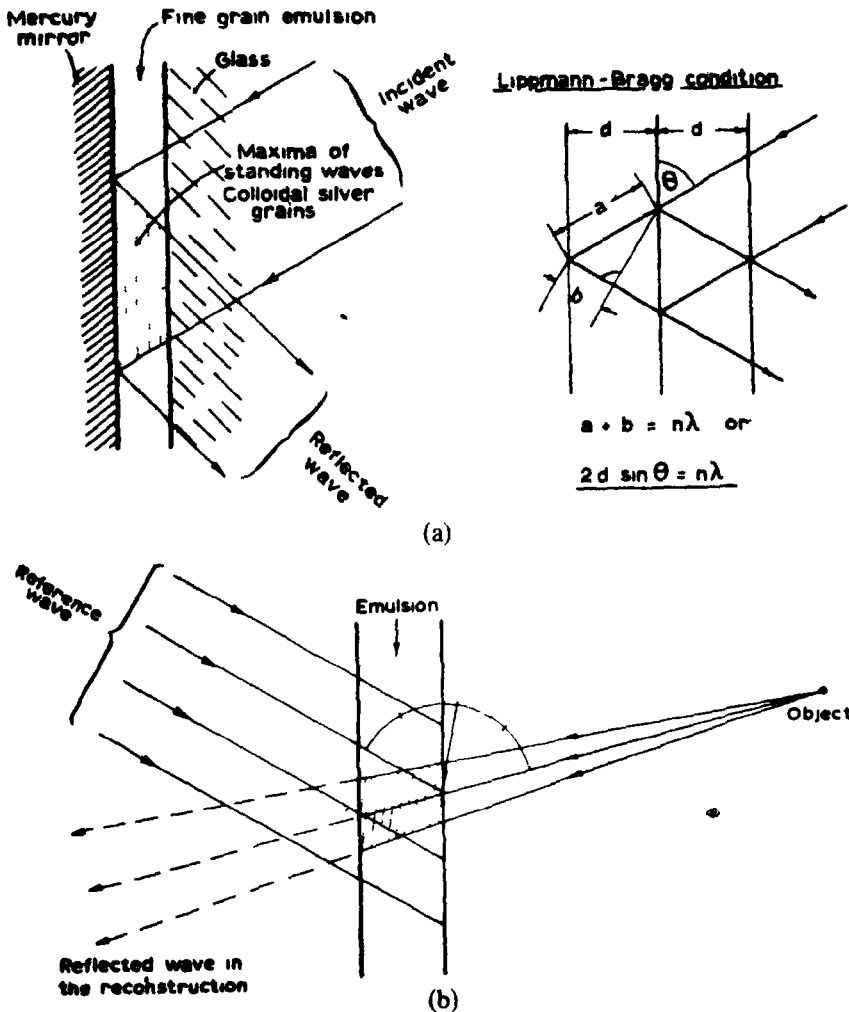



Fig 26 (a) Gabriel Lippmann's method of photography in natural colours (b) Lippmann-Denisyuk-Stroke reflection hologram

Fig. 27. First two color reflecting hologram reconstructed in white light [L. H. Lin, K. S. Pennington, G. W. Stroke and A. E. Labeyrie (14)]

but two-, and even more, three-colour holograms are still far from being satisfactory. It is one of my chief preoccupations at the present to improve this situation, but it would take too long, and it would be also rather early to enlarge on this.

An application of holography which is certain to gain high importance in the next years is information storage. I have mentioned before that holography allows storing 100–300 times more printed pages in a given emulsion than ordinary microphotography. Even without utilizing the depth dimension, the factor is better than 50. The reason is that a diffused hologram represents almost ideal coding, with full utilization of the area and of the gradation of the emulsion, while printed matter uses only about 5–10% of the area, and the gradation not at all. A further factor arises from the utilization of the third dimension, the depth of the emulsion. This possibility was first pointed out in an ingenious paper by P. J. van Heerden [16], in 1963. Theoretically it appears possible to store one bit of information in about one wavelength cube. This is far from being practical, but the figure of 300, previously mentioned, is entirely realistic.

However, even without this enormous factor, holographic storage offers important advantages A binary store, in the form of a checker-board pattern on microfilm can be spoiled by a single grain of dust, by a hair or by a scratch, while a diffused hologram is almost insensitive to such defects The holographic store, illustrated in Fig 28, is according to its author L K Anderson [17] (1968) only a modest beginning, yet it is capable of accessing for instance any one of 64 × 64 printed pages in about a microsecond Each hologram, with a diameter of 1 2 mm can contain about 10^4 bits Reading out this information sequentially in a microsecond would of course require an impossible waveband, but powerful parallel reading means can be provided One can confidently expect enormous extensions of these "modest beginnings" once the project of data banks will be tackled seriously

Another application of holography, which is probably only in an early stage, is pattern and character recognition I can only briefly refer to the basic work which A Vander Lugt [18] has done in the field of pattern recognition It will be sufficient to explain the basic principle of character recognition with the aid of Fig 29

Let us generalize a little the basic principle of holography In all previous examples a complicated object beam was brought to interference with a simple plane or spherical reference beam, and the object beam was reconstructed by illuminating the hologram with the reference beam But a little mathematics shows that this can be extended to any reference beam *which correlates sharply with itself* The autocorrelation function is an invariant of a beam, it can be computed in any cross section One can see at once that a spherical wave correlates sharply with itself, because it issues from a "point " But there are other beams which correlate sharply with themselves, for instance those which issue from a fingerprint, or from

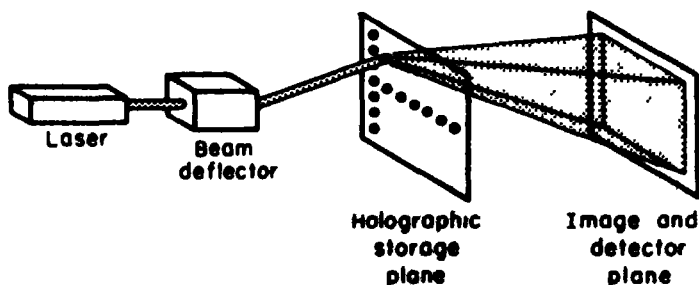



Fig. 28 Holographic flying spot store L K Anderson and R J Collier, Bell Telephone Laboratories, 1968

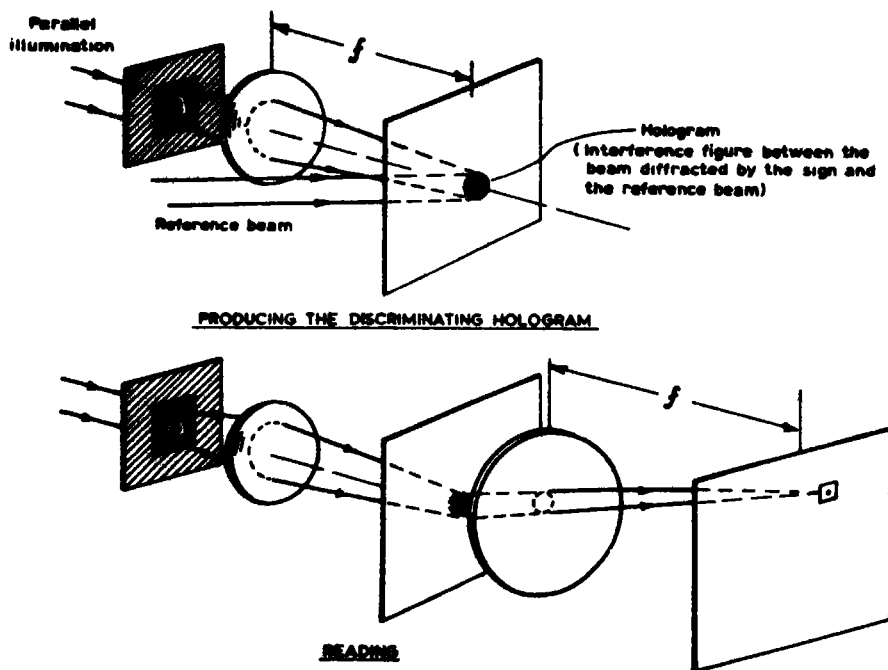



Fig 29 The principle of character recognition by holography

a Chinese ideogram, in an extreme case also those which issue from a piece of frosted glass Hence it is quite possible for instance to translate, by means of a hologram, a Chinese ideogram into its corresponding English sentence and *vice versa* J N Butters and M Wall in Loughborough University have recently created holograms which from a portrait produce the signature of the owner, and *vice versa* In other words, a hologram can be a fairly universal translator It can for instance translate a sign which we can read to another which a machine can read

Figure 29 shows a fairly modest realisation of this principle A hologram is made of a letter "a" by means of a plane reference beam When this hologram is illuminated with the letter "a" the reference beam is reconstructed, and can activate for instance a small photocell in a certain position This, I believe, gives an idea of the basic principle There is of course much more needed to make a practical system of it, because there are so many ways of printing letters, but it would take me too long to explain how to deal with this and other difficulties

With character recognition devices we have already taken half a step

into the future, because these are likely to become important only in the next generation of computers or robots, to whom we must transfer a little more of human intelligence I now want to mention briefly some other problems which are half or more than half in the future

One, which is already very actual, is the overcoming of laser speckle Everybody who sees laser light for the first time is surprised by the rough appearance of objects which we consider as smooth A white sheet of paper appears as if it were crawling with ants The crawling is put into it by the restless eye, but the roughness is real It is called "laser speckle" and Fig 30 shows a characteristic example of it This is the appearance of

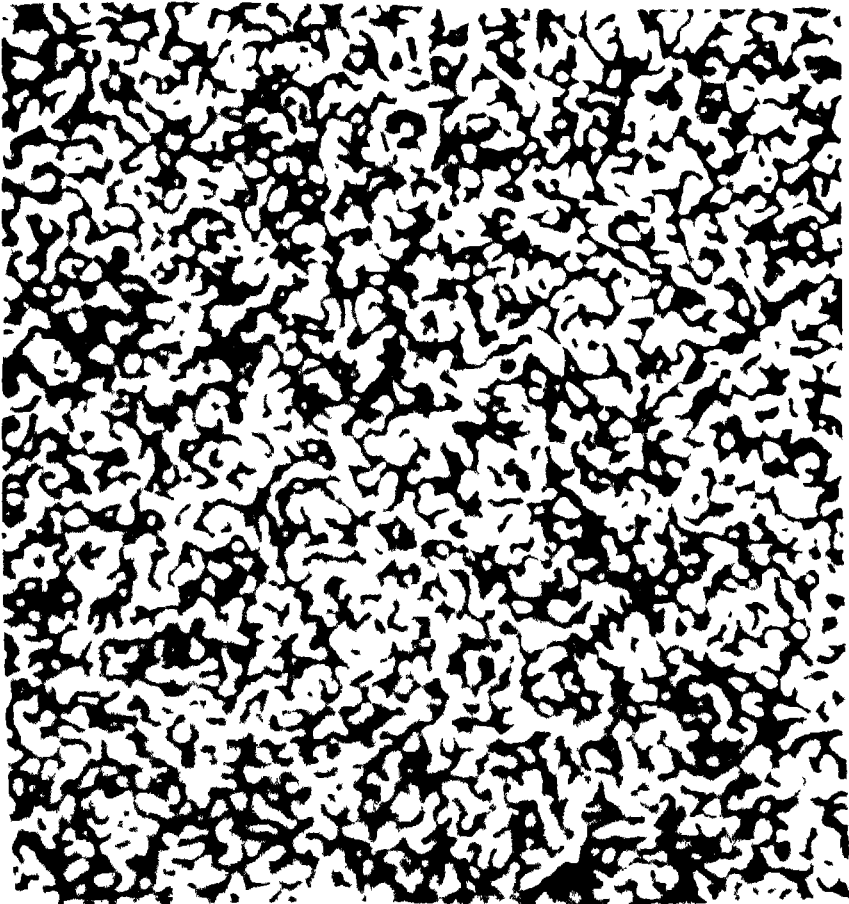



Fig 30 Laser speckle The appearance of e g a white sheet of paper, uniformly illuminated by laser light

a white sheet of paper in laser light, when viewed with a low-power optical system It is not really noise, it is information which we do not want, information on the microscopic unevenness of the paper in which we are not interested What can we do against it?

In the case of rough objects the answer is, regrettably, that all we can do is to average over larger areas, thus smoothing the deviations This means that we must throw a great part of the information away, the wanted with the unwanted This is regrettable but we can do nothing else, and in many cases we have enough information to throw away, as can be seen by the fully satisfactory appearance of some of the reconstructions from diffuse holograms which I have shown However, there are important areas in which we can do much more, and where an improvement is badly needed This is the area of microholograms, for storing and for display They are made as diffused holograms, in order to ensure freedom from dust and scratches, but by making them diffused, we introduce speckle, and to avoid this such holograms are made nowadays much larger than would be ideally necessary I have shown recently [19], that advantages of diffuse holograms can be almost completely retained, while the speckle can be completely eliminated by using, instead of a frosted glass, a special illuminating system This, I hope will produce a further improvement in the information density of holographic stores

Now let us take a more radical step into the future I want to mention briefly two of my favourite holographic brainchilden The first of this is Panoramic Holography, or one could also call it Holographic Art

All the three-dimensional holograms made so far extend to a depth of a few meters only Would it not be possible to extend them to infinite? Could one not put a hologram on the wall, which is like a window through which one looks at a landscape, real or imaginary? I think it can be done, only it will not be a photograph but a work of art Figure 31 illustrates the process The artist makes a model, distorted in such a way that it appears perspectivic, and extending to any distance when viewed through a large lens, as large as the hologram The artist can use a smaller lens, just large enough to cover both his eyes when making the model A reflecting hologram is made of it, and illuminated with a strong, small light source The viewer will see what the plate has seen through the lens, that is to say a scene extending to any distance, in natural colours This scheme is under development, but considerable work will be needed to make it satisfactory, because we must first greatly improve the reflectance of three-colour holograms

An even more ambitious scheme, probably even farther in the future, is three-dimensional cinematography, without viewing aids such as Polaroids The problem is sketched out in Fig 32 The audience (in one

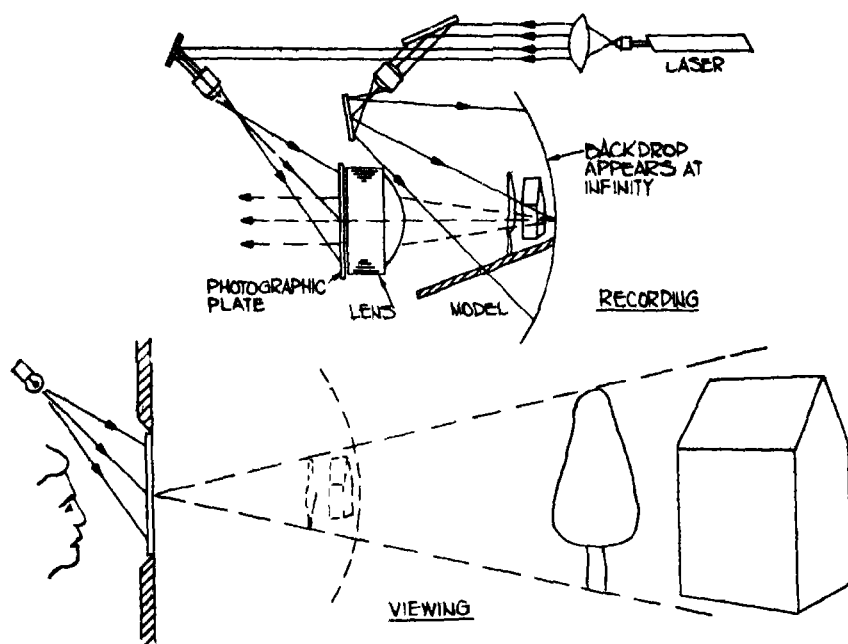



Fig 31 Panoramic holography

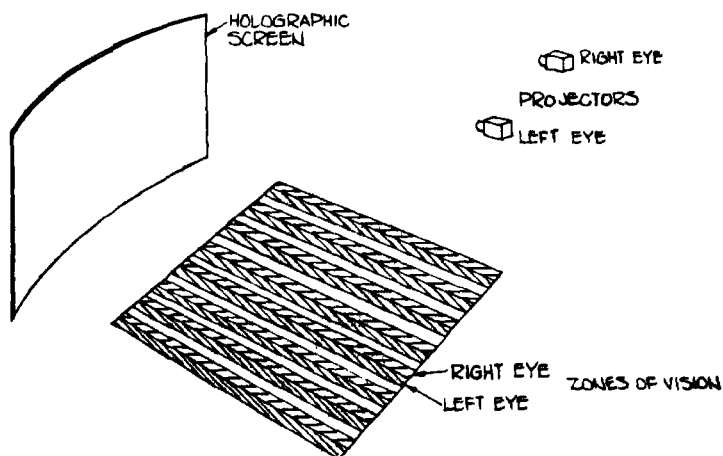



Fig 32 3-D Cinematography with holographic screen

plane or two) is covered by zones of vision, with the width of the normal eye spacing, one for the right eye, one for the left, with a blank space between two pairs The two eyes must see two different pictures, a stereoscopic pair The viewer can move his head somewhat to the right or left Even when he moves one eye into the blank zone, the picture will appear dimmer but not flat, because one eye gives the impression of "stereoscopy by default "

I have spent some years of work on this problem, just before holography, until I had to realise that it is strictly unsolvable with the orthodox means of optics, lenticules, mirrors, prisms One can make satisfactorily small screens for small theatres, but with large screens and large theatres one falls into a dilemma If the lenticules, or the like, are large, they will be seen from the front seats, if they are small, they will not have enough definition for the back seats

Some years ago I realised to my surprise, that holography can solve this problem too Use a projector as the reference source, and for instance the system of left viewing zones as the object The screen, covered with a Lippmann emulsion, will then make itself automatically into a very complicated optical system such that when a picture is projected from the projector, it will be seen only from the left viewing zones One then repeats the process with the right projector, and the right viewing zones Volume (Lippmann-Denisyuk) holograms display the phenomenon of directional selectivity If one displaces the illuminator from the original position by a certain angle, there will be no reflection We put the two projectors at this angle (or a little more) from one another, and the effect is that the right picture will not be seen by the left eye and vice versa

There remains of course one difficulty, and this is that one cannot practice holography on the scale of a theatre, and with a plate as large as a screen But this too can be solved, by making up the screen from small pieces, not with the theatre but with a *model* of the theatre, seen through a lens, quite similar to the one used in panoramic holography

I hope I have conveyed the feasibility of the scheme, but I feel sure that I have conveyed also its difficulties I am not sure whether they will be overcome in this century, or in the next

Ambitious schemes, for which I have a congenital inclination, take a long time for their realisation As I said at the beginning, I shall be lucky if I shall be able to see in my lifetime the realisation of holographic electron microscopy, on which I have started 24 years ago But I have good hope, because I have been greatly encouraged by a remarkable achievement of G W Stroke [20], which is illustrated in Figure 33 Professor Stroke has recently succeeded in deblurring micrographs taken

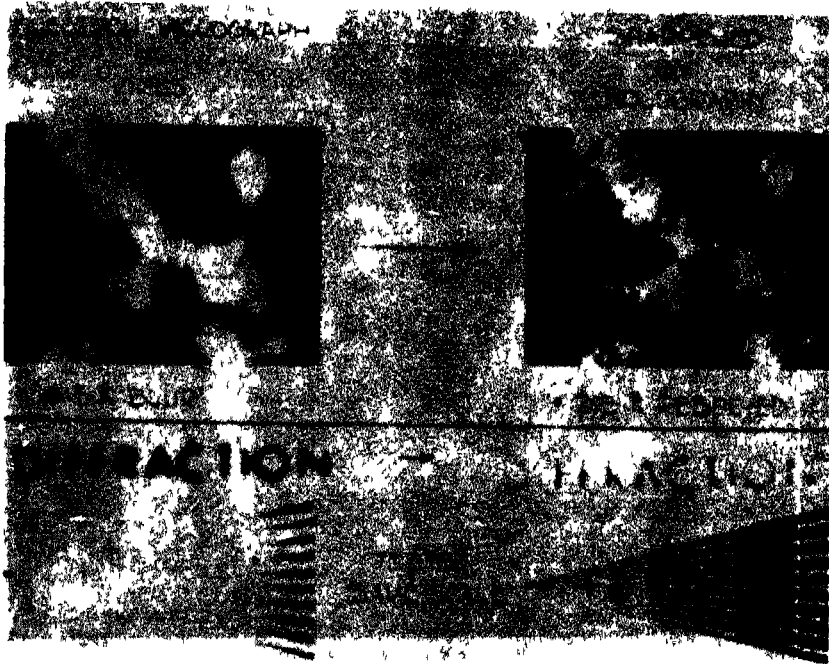



Fig 33 Scanning transmission electron micrograph Professor Albert Crewe, University of Chicago holographically deblurred by Professor G W Stroke 1971 The bottom photographs prove that the effect could not be obtained by hard printing because some spatial frequencies which appear in the original with reversed phase had to be phase-corrected

by Professor Albert Crewe, Chicago, with his scanning transmission electron microscope, by a holographic filtering process, improving the resolution from 5 Ångstrom to an estimated 2 5 Ångstrom This is not exactly holographic electron microscopy, because the original was not taken with coherent electrons, but the techniques used by both sides, by A Crewe and by G W Stroke are so powerful, that I trust them to succeed also in the next, much greater and more important step

Summing up, I am one of the few lucky physicists who could see an idea of theirs grow into a sizeable chapter of physics I am deeply aware that this has been achieved by an army of young, talented and enthusiastic researchers, of whom I could mention only a few by name I want to express my heartfelt thanks to them, for having helped me by their work to this greatest of scientific honours

Holographic portrait of Dennis Gabor See page 383 for explanation

Bibliography and References

It is impossible to do justice to the hundreds of authors who have significantly contributed to the development of holography The number of articles exceeds 2000, and there are more than a dozen books in several languages

An extensive bibliography may be found for instance in T Kallard (ed), *Holography*, Optosonic Press, New York (1969 and 1970)

Books

Barrekette E S Kock, W E, Ose, T, Tsujiuchi, J, and Stroke, G W (eds) (1971), *Applications of Holography*, Plenum Press New York

Caulfield, H J and Sun Lu (1970) *The Applications of Holography* Wiley Interscience, New York

Collier, R J, Burckhardt, C B, and Lin, L H (1971), *Optical Holography*, Academic Press, New York

DeVelis J B, and Reynolds, G O (1967), *Theory and Applications of Holography*, Addision Wesley, Reading, Massachusetts

Françon, M, (1969), *Holographie* Masson et Cie Paris

Kiemle, H, and Ross, D (1969), *Einfuhrung in die Technik der Holographie*, Akademische Verlagsgesellschaft, Frankfurt am Main, transl (1972), *Introduction to Holographic Techniques*, Plenum Press, 1972

Kock, W E (1969), *Lasers and Holography* (*An Introduction to Coherent Optics*) Doubleday & Co, Garden City, New York

Ostrovsky Yu I (1970), *Holography* (in Russian) Nauka Leningrad

Robertson, E R, and Harvey J M (eds) (1970) *The Engineering Uses of Holography*, Cambridge University Press, Cambridge

Stroke, G W (1966) *An Introduction to Coherent Optics and Holography*, Academic Press, New York (second edition 1969)

Vienot, J Ch, Smigielski P and Royer H (1971) *Holographie Optique* (*Developpements, Applications*), Dunod Paris

Papers

1 D Gabor, A new microscopic principle, *Nature* **161**, No 4098 777–778 (1948), Microscopy by reconstructed wavefronts *Proc R Soc London Ser* A **197**, 454-487 (1949), Microscopy by reconstructed wavefronts II, *Proc Roy Soc* (*London*) **64** (Pt 6) No 378 B 449–469 (1951)

2 G L Rogers, Experiments in diffraction microscopy *Proc R Soc Edinburgh* **63A**, 193 (1952)

3 A Baez, Resolving power in diffraction microscopy, *Nature* **169**, 963–964 (1952)

4 H M A El-Sum and P Kirkpatrick Microscopy by reconstructed wavefronts, *Phys Rev* **85**, 763 (1952)

5 D Gabor and W P Goss Interference microscope with total wavefront reconstruction *J Opt Soc Am* **56**, 849–858 (1966)

6 L J Cutrona E N Leith, L J Porcello and W E Vivian On the application of coherent optical processing techniques to synthetic aperture radar *Proc IEEE* **54**, 1026–1032 (1966)

7 E N Leith and J Upatnieks Wavefront reconstruction with continuous tone transparencies *J Opt Soc Am* **53**, 522 (1963) (Abstract) Wavefront reconstruction with continuous tone objects *J Opt Soc Am* **53**, 1377–1381 (1963) Wavefront reconstruction with diffused illumination and three-dimensional objects, *J Opt Soc Am* **54**, 1295 (1964)

8 An early reference is G W Stroke Theoretical and experimental foundations of high-resolution optical holography (Presented in Rome on 14 September 1964), in Vol II, pp 53–63 of *Pubblicazioni IV Centenario della Nascita di Galileo Galilei* G Barbera, Firenze (1966)

9 C Knox and R E Brooks, Holographic motion picture microscopy *Proc R Soc London Ser* B **174**, 115–121 (1969)

10 L O Heflinger, R F Wuerker, and R E Brooks *J Appl Phys* **37**, No 2, 642–649 (February 1966)

11 J M Burch, The application of lasers in production engineering, *Prod Eng* (*London*) **44**, 431–442 (1965) R L Powell and K A Stetson Interferometric vibration analysis by wavefront reconstruction *J Opt Soc Am* **55**, 1593–1598 (1965), G W Stroke and

A. E. Labeyrie, Two-beam interferometry by successive recording of intensities in a single hologram, *Appl. Phys. Lett.* **8**, No 2, 42–44 (15 January 1966)
12 K. A. Stetson, thesis (under direction of E Ingelstam), Royal Institute of Technology, Stockholm (1969)
13 Yu. N Denisyuk, Photographic reconstruction of the optical properties of an object in its own scattered radiation, *Dokl. Akad Nauk SSSR*, **144**, 1275–1278 (1962)
14a G. W Stroke and A E Labeyrie, White-light reconstruction of holographic images using the Lippmann–Bragg diffraction effect, *Phys Lett* **20**, 368–370 (1966)
14b L. H Lin, K S Pennington, G W Stroke, and A E Labeyrie, *Bell Syst Tech J*, **45**, 659 (1966)
15 K. S Pennington and J S Harper, Techniques for producing low-noise, improved-efficiency holograms, *Appl. Opt.* **9**, 1643–1650 (1970)
16 P J Van Heerden, A new method of storing and retrieving information, *Appl Opt* **2**, 387–392 (1963)
17 L. K Anderson, Holographic optical memory for bulk data storage, *Bell Lab Rep* **46**, 318 (1968)
18 A Vander Lugt, "Signal Detection by Complex Spatial Filtering", *IEEE Trans Inf Theory* **IT—10**, 139–145 (1964)
19 D Gabor, Laser speckle and its elimination, *IBM J Res Dev* **14**, 509–514 (Sept 1970)
20 G W Stroke, Image deblurring and aperture synthesis using *a posteriori* processing by Fourier-transform holography, *Opt. Act* **16**, 401–422 (1971), G W Stroke, and M Halioua, Attainment of diffraction-limited imaging in high-resolution electron microscopy by *a posteriori* holographic image sharpening, *Optik* (1972)

Appendix

A The Fourier Transform

According to Fourier's theorem, a periodic function can be expressed as a sum of sine and cosine functions whose frequencies increase in the ratio of natural numbers Thus, a periodic function with period α, i e ,

$$f(x + n\alpha) = f(x), \qquad n = 0 \pm 1, \pm 2, \pm 3, \tag{A-1}$$

expanded in the form of

$$f(x) = \tfrac{1}{2}a_0 + \sum_{n=1}^{\infty} [a_n \cos(nkx) + b_n \sin(nkx)] \tag{A-2}$$

where

$$k = \frac{2\pi}{\alpha} \tag{A-3}$$

The coefficients a_n and b_n can easily be determined by using the following properties of the trigonometric functions

$$\frac{2}{\alpha}\int_{x_0}^{x_0+\alpha} \cos(nkx)\cos(mkx)\,dx = \delta_{mn} \tag{A-4}$$

$$\frac{2}{\alpha}\int_{x_0}^{x_0+\alpha} \sin(nkx)\sin(mkx)\,dx = \delta_{mn} \tag{A-5}$$

$$\frac{2}{\alpha}\int_{x_0}^{x_0+\alpha} \sin(nkx)\cos(mkx)\,dx = 0 \tag{A-6}$$

where x_0 is arbitrary and δ_{mn} is the Kronecker delta function defined through the following equation

$$\delta_{mn} = \begin{cases} 0 & \text{if} \quad m \neq n \\ 1 & \text{if} \quad m = n \end{cases} \tag{A-7}$$

If we multiply Eq (A-1) by $\cos(mkx)$ and integrate from x_0 to $x_0 + \alpha$, we would obtain

$$a_n = \frac{2}{\alpha} \int_{x_0}^{x_0+\alpha} f(x) \cos(nkx)\, dx, \qquad n = 0, 1, 2, \tag{A-8}$$

Similarly

$$b_n = \frac{2}{\alpha} \int_{x_0}^{x_0+\alpha} f(x) \sin(nkx)\, dx, \qquad n = 1, 2, 3, \tag{A-9}$$

For the sake of convenience we choose $x_0 = -\alpha/2$ If we now substitute the above expressions for a_n and b_n in Eq (A-2) we would obtain

$$f(x) = \frac{1}{\alpha} \int_{-\alpha/2}^{\alpha/2} f(x')\, dx' + \sum_{n=1}^{\infty} \left[\frac{2}{\alpha} \cos(nkx) \int_{-\alpha/2}^{+\alpha/2} f(x') \cos(nkx')\, dx' + \frac{2}{\alpha} \sin(nkx) \int_{-\alpha/2}^{+\alpha/2} f(x') \sin(nkx')\, dx' \right] \tag{A-10}$$

or

$$f(x) = \frac{1}{2\pi} \Delta s \int_{-\pi/\Delta s}^{+\pi/\Delta s} f(x')\, dx' + \sum_{n=1}^{\infty} \left\{ \frac{\Delta s}{\pi} \int_{-\pi/\Delta s}^{+\pi/\Delta s} f(x') \cos[n\, \Delta s (x' - x)]\, dx' \right\} \tag{A-11}$$

where

$$\Delta s = \frac{2\pi}{\alpha} = k \tag{A-12}$$

We now let $\alpha \to \infty$ so that $\Delta s \to 0$ Thus if the integral

$$\int_{-\infty}^{+\infty} |f(x')|\, dx'$$

exists (i e , it has a finite value), then the first term on the right-hand side of Eq (A-11) would go to zero Further, since

$$\int_0^{\infty} F(s)\, ds = \lim_{\Delta s \to 0} \sum_{n=1}^{\infty} F(n\, \Delta s)\, \Delta s \tag{A-13}$$

we have, in the limit $\Delta s \to 0$,

$$f(x) = \frac{1}{\pi} \int_0^{\infty} \left\{ \int_{-\infty}^{+\infty} f(x') \cos[s(x' - x)]\, dx' \right\} ds \tag{A-14}$$

Equation (A-14) is known as the Fourier integral Since the cosine function inside the integral is an even function of s, we may write

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \left\{ \int_{-\infty}^{+\infty} f(x') \cos[s(x' - x)]\, dx' \right\} ds \tag{A-15}$$

Further, since $\sin[s(x' - x)]$ is an odd function of s

$$0 = \frac{i}{2\pi} \int_{-\infty}^{+\infty} \left\{ \int_{-\infty}^{+\infty} f(x') \sin[s(x' - x)]\, dx' \right\} ds \tag{A-16}$$

Adding (or subtracting) Eqs (A-15) and (A-16), we get

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x') e^{\pm is(x' - x)}\, dx'\, ds \tag{A-17}$$

Thus, if

$$F(s) = \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{\infty} f(x') e^{isx}\, dx' \tag{A-18}$$

then

$$f(x) = \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{+\infty} F(s) e^{-isx}\, ds \tag{A-19}$$

The function $F(s)$ defined by Eq (A-18) is known as the Fourier transform of $f(x)$ and conversely

Since the Dirac delta function $\delta(x - x')$ is defined by the equation

$$f(x) = \int_{-\infty}^{+\infty} f(x') \delta(x - x')\, dx' \tag{A-20}$$

we obtain, comparing Eqs (A-17) and (A-20), the following representation of the delta function

$$\delta(x' - x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{\pm is(x' - x)}\, ds \tag{A-21}$$

For a time-dependent function, we can write the Fourier transform in the following form

$$F(\omega) = \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{+\infty} f(t') e^{i\omega t}\, dt' \tag{A-22}$$

$$f(t) = \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{+\infty} F(\omega) e^{-i\omega t}\, d\omega \tag{A-23}$$

where ω represents the angular frequency As an example, we consider a time dependent pulse of the form

$$f(t) = A e^{-t^2/2\tau^2} e^{-i\omega_0 t} \tag{A-24}$$

Thus,

$$F(\omega) = \frac{A}{(2\pi)^{1/2}} \int_{-\infty}^{+\infty} \exp\left(-\frac{t^2}{2\tau^2}\right) e^{i(\omega-\omega_0)t}\, dt$$

$$= A\tau \exp\left[-\frac{(\omega-\omega_0)^2\tau^2}{2}\right] \tag{A-25}$$

where we have used the following result

$$\int_{-\infty}^{+\infty} e^{-\alpha x^2+\beta x}\, dx = e^{\beta^2/4\alpha} \int_{-\infty}^{+\infty} \exp\left[-\alpha\left(x - \frac{\beta}{2\alpha}\right)^2\right] dx$$

$$= (\pi/\alpha)^{1/2} \exp(\beta^2/4\alpha) \tag{A-26}$$

From Eqs (A-24) and (A-25) it can immediately be seen that a temporal pulse of duration $\sim \tau$ has a frequency spread $\Delta\omega \sim 1/\tau$ Thus, one obtains

$$\tau\, \Delta\omega \sim 1 \tag{A-27}$$

If $\tau \to \infty$, the pulse becomes almost monochromatic and $\Delta\omega \to 0$ The result expressed by Eq (A-27) is quite general in the sense that it is independent of the shape of the pulse For example, for a pulse of the form

$$f(t) = \begin{cases} Ae^{-i\omega_0 t}, & |t| < \tau/2 \\ 0 & \text{everywhere else,} \end{cases} \tag{A-28}$$

one would obtain

$$F(\omega) = A\left(\frac{2}{\pi}\right)^{1/2} \frac{\sin(\omega - \omega_0)\tau/2}{\omega - \omega_0} \tag{A-29}$$

which also has a spread $\sim 1/\tau$

If we introduce the variable

$$u = s/2\pi \tag{A-30}$$

then Eq (A-17) becomes

$$f(x) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x')e^{2\pi iu(x'-x)}\, dx'\, du \tag{A-31}$$

Thus, if we write

$$F(u) = \int_{-\infty}^{+\infty} f(x')e^{2\pi iux'}\, dx' = \mathscr{F}[f(x)] \tag{A-32}$$

then

$$f(x) = \int_{-\infty}^{+\infty} F(u)e^{-2\pi iux}\, du \tag{A-33}$$

where the symbol $\mathscr{F}[\ ]$ stands for the Fourier transform of the quantity inside the brackets The quantity u is termed the spatial frequency (see Section 10 2)

Now

$$\begin{aligned}\mathscr{F}[\mathscr{F}[f(x)]] &= \mathscr{F}[F(u)] \\ &= \int_{-\infty}^{+\infty} F(u)e^{2\pi i u x}\,dx \\ &= f(-x) \qquad \text{(A-34)}\end{aligned}$$

Thus the Fourier transform of the Fourier transform of a function is the original function itself except for an inversion We also see that

$$\begin{aligned}\mathscr{F}\left[\exp\left(\frac{2\pi i x}{a}\right)\right] &= \int_{-\infty}^{+\infty} \exp\left[2\pi i x\left(\frac{1}{a} + u\right)\right] dx \\ &= \delta(u + 1/a) \qquad \text{(A-35)}\end{aligned}$$

where we have used Eq (A-21) In addition,

$$\begin{aligned}\mathscr{F}[\delta(t - t_0)] &= \int_{-\infty}^{\infty} \delta(t - t_0)e^{2\pi i \nu t}\,dt \\ &= e^{2\pi i \nu t_0} \qquad \text{(A-36)}\end{aligned}$$

Convolution Theorem The convolution of two functions $f(t)$ and $g(t)$ is defined by the relation

$$f(t) * g(t) = \int_{-\infty}^{\infty} f(t')g(t - t')\,dt' = g(t) * f(t) \qquad \text{(A-37)}$$

We will now show that the Fourier transform of the convolution of two functions is the product of their Fourier transforms This can be seen as follows

$$\begin{aligned}\mathscr{F}[f(t) * g(t)] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(t')g(t - t')e^{2\pi i \nu t}\,dt'\,dt \\ &= \int_{-\infty}^{\infty} dt'\, f(t')e^{2\pi i \nu t'} \int_{-\infty}^{\infty} g(t - t')e^{2\pi i \nu(t-t')}\,dt \\ &= F(\nu)G(\nu) \qquad \text{(A-38)}\end{aligned}$$

where $F(\nu) = \mathcal{F}[f(t)]$ and $G(\nu) = \mathcal{F}[g(t)]$ Similarly

$$
\begin{aligned}
\mathcal{F}[f(t)g(t)] &= \int_{-\infty}^{\infty} f(t')g(t')e^{+2\pi i\nu t'}\,dt' \\
&= \int_{-\infty}^{\infty} f(t') \int_{-\infty}^{\infty} G(\nu')e^{-2\pi i\nu' t'}\,d\nu'\,e^{2\pi i\nu t'}\,dt' \\
&= \int_{-\infty}^{\infty} G(\nu') \int_{-\infty}^{\infty} f(t')e^{2\pi i(\nu-\nu')t'}\,dt'\,d\nu' \\
&= \int_{-\infty}^{\infty} G(\nu')F(\nu - \nu')\,d\nu' \\
&= F(\nu) * G(\nu) \qquad \text{(A-39)}
\end{aligned}
$$

i e , the Fourier transform of the product of two functions is the convolution of their Fourier transforms

We will now consider some examples The convolution of a function $f(t)$ with a delta function $\delta(t - t_0)$ is

$$
\begin{aligned}
f(t) * \delta(t - t_0) &= \int_{-\infty}^{\infty} f(t')\delta(t - t' - t_0)\,dt' \\
&= f(t - t_0) \qquad \text{(A-40)}
\end{aligned}
$$

Thus the convolution of a function with a delta function yields the same function but with a shift in origin

Similarly, the convolution of two Gaussian functions can again be shown to be another Gaussian function

$$
\exp\left(-\frac{x^2}{\alpha^2}\right) * \exp\left(-\frac{x^2}{\beta^2}\right) = \frac{\alpha\beta(\pi)^{1/2}}{(\alpha^2 + \beta^2)^{1/2}}\exp\left(-\frac{x^2}{\alpha^2 + \beta^2}\right) \qquad \text{(A-41)}
$$

Sampling Theorem We will now derive the sampling theorem according to which *a band limited function, i e , a function which has no spectral components beyond a certain frequency, say* ν_m *is uniquely determined by its value at uniform intervals less than* $(1/2\nu_m)$ *seconds*

A band limited function having a maximum spectral component ν_m implies that its Fourier transform is zero beyond ν_m (see Fig A1) Thus if $f(t)$ is a band limited function, then

$$
\mathcal{F}[f(t)] = F(\nu) = 0 \qquad \text{for} \quad \nu > \nu_m \qquad \text{(A-42)}
$$

We now consider the sampled function which consists of impulses every T seconds with the strength of each impulse being equal to the value of the function at that time Thus the sampled function can be

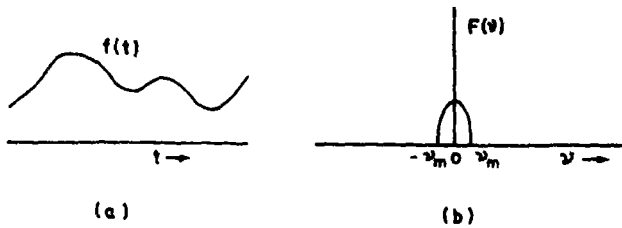



Fig A 1 (a) A band-limited function having no spectral components beyond a frequency ν_m and (b) the corresponding Fourier spectrum

written as

$$f_s(t) = \sum_{n=-\infty}^{\infty} f(t)\,\delta(t - nT) = f(t) \sum_{n=-\infty}^{\infty} \delta(t - nT) \qquad \text{(A-43)}$$

The sampled function $f_s(t)$ is shown in Fig A 2a Here T represents the sampling interval and $\delta(t - nT)$ represents the Dirac delta function

In order to obtain the Fourier spectrum of the sampled function, we have to first find the Fourier transform of the function $\sum_{n=-\infty}^{\infty} \delta(t - nT)$ Using Eq (A-36), we have

$$\mathscr{F}\left[\sum_{n=-\infty}^{\infty} \delta(t - nT)\right] = \int_{-\infty}^{\infty} \sum_{n=-\infty}^{\infty} \delta(t - nT) e^{2\pi i\nu t}\, dt$$

$$= \sum_{n=-\infty}^{\infty} e^{2\pi i n\nu T} \qquad \text{(A-44)}$$

In order to evaluate the sum on the right-hand side of Eq (A-44), we

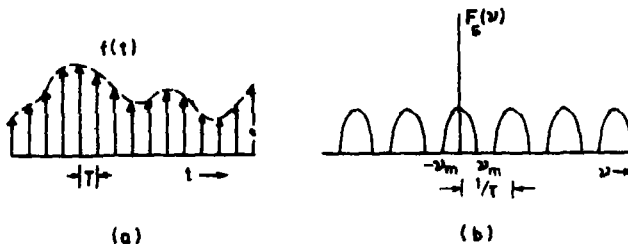



Fig A 2 (a) The function shown here is obtained by periodically sampling the original function depicted in Fig A 1 at uniform intervals by impulses separated in time by T, the strength of the sampled signal at the sampling times is equal to the value of the original function at that time (b) The Fourier spectrum corresponding to the sampled function when the sampling interval T is smaller than $1/2\nu_m$, the Nyquist interval Observe that in such a case the various repetitively appearing spectra do not overlap and it is possible to use a filter to retrieve $F(\nu)$ and hence $f(t)$

consider the following finite sum

$$\sum_{n=-N}^{N} e^{2\pi i\nu nT} = e^{-2\pi i\nu NT}\left(\frac{1-e^{2\pi i\nu T(2N+1)}}{1-e^{2\pi i\nu T}}\right) = \frac{\sin[\pi\nu T(2N+1)]}{\sin(\pi\nu T)} \quad \text{(A-45)}$$

which is a periodic function with a period $1/T$ If we restrict ourselves to the region $-1/2T < \nu < 1/2T$, then for a large value of N, the above function is sharply peaked around $\nu = 0$ and we may write

$$\lim_{N\to\infty} \frac{\sin \pi\nu T(2N+1)}{\sin \pi\nu T} = \lim_{g\to\infty} \frac{\sin g\nu T}{\pi\nu T} \quad \text{(A-46)}$$

In order to determine the value of the limit on the right-hand side of Eq (A-46), we have from Eq (A-21)

$$\begin{aligned}\delta(t) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{i\nu t}\, d\nu \\ &= \lim_{N\to\infty} \frac{1}{2\pi}\int_{-N}^{N} e^{i\nu t}\, d\nu \\ &= \lim_{N\to\infty} \frac{\sin Nt}{\pi t} \end{aligned} \quad \text{(A-47)}$$

Thus we get from Eq (A-46)

$$\lim_{N\to\infty} \frac{\sin \pi\nu T(2N+1)}{\sin \pi\nu T} = \delta(\nu T) \quad \text{(A-48)}$$

and from Eqs (A-44) and (A-45)

$$\mathscr{F}\left[\sum_{n=-\infty}^{\infty} \delta(t-nT)\right] = \frac{1}{T}\sum_{n=-\infty}^{\infty} \delta\left(\nu - \frac{n}{T}\right) \quad \text{(A-49)}$$

Hence the Fourier transform of the sampled function is given by

$$\begin{aligned}\mathscr{F}[f_s(t)] &= \mathscr{F}\left[f(t)\sum_{n=-\infty}^{\infty} \delta(t-nT)\right] \\ &= \mathscr{F}[f(t)] * \mathscr{F}\left[\sum_{n=-\infty}^{\infty} \delta(t-nT)\right] \quad \text{using Eq (A-39)} \\ &= F(\nu) * \frac{1}{T}\sum_{n=-\infty}^{\infty} \delta\left(\nu - \frac{n}{T}\right) \quad \text{using Eq (A-49)} \\ &= \frac{1}{T}\sum_{n=-\infty}^{\infty} F(\nu) * \delta\left(\nu - \frac{n}{T}\right) \\ &= \frac{1}{T}\sum_{n=-\infty}^{\infty} F\left(\nu - \frac{n}{T}\right) \quad \text{using Eq (A-40)} \end{aligned} \quad \text{(A-50)}$$

Here

$$F(\nu) = \mathcal{F}[f(t)] = \int_{-\infty}^{\infty} f(t)e^{2\pi i\nu t}\,dt \tag{A-51}$$

represents the Fourier transform of the function $f(t)$ Hence from above it follows that the Fourier transform of the sampled function consists of an infinite sum of repetitively appearing Fourier transforms of the original function (see Fig A 2b), i e, the function $F(\nu)$ appears at uniform intervals of $1/T$ Thus, if the function $f(t)$ is band limited up to ν_m, then in order that the various repetitively appearing Fourier spectra in the sampled function do not overlap, one must have

$$2\nu_m \leq 1/T$$

or

$$T \leq 1/2\nu_m \tag{A-52}$$

Hence as long as we sample $f(t)$ at regular intervals less than $1/2\nu_m$ seconds apart, the Fourier spectrum of the sampled function is a periodic replica of $F(\nu)$ One can thus recover $f(t)$ completely by allowing the sampled signal to pass through a low-pass filter which attenuates all frequencies beyond ν_m and which passes without distortion the frequency components below ν_m On the other hand, if $T > 1/2\nu_m$, then the various spectra $F(\nu)$ overlap and it is not possible to recover $f(t)$ from the sampled values (see Fig A 3) Thus in order to completely preserve the information content of a signal in a retrievable form, the sampling interval should not be more than $1/2\nu_m$, this maximum sampling interval is also referred to as the Nyquist interval

Hence from the above theorem it follows that complete information of a band limited signal can be transmitted by just sending the discrete sampled values This is the basic principle behind digital transmission systems (see Chapter 12)

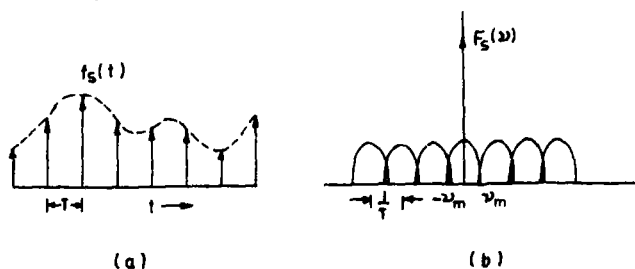



Fig A 3 (a) The same as shown in Fig A 2a but here the sampling interval is larger than $1/2\nu_m$ (b) The corresponding Fourier spectrum which shows overlap of the various repetitively appearing spectra In such a case it is not possible to retrieve the signal

It may be mentioned here that no signal is strictly band limited But for most practical situations the energy content in high frequencies is so small as to be negligible Thus, one can consider such signals to be essentially band limited

B. Propagation of a Gaussian Wave Packet

For a one-dimensional Gaussian wave packet [see Eq (2 2-12)]

$$\Psi(x, t = 0) = \frac{1}{(\pi\sigma^2)^{1/4}} \exp\left[-\frac{(x - x_0)^2}{2\sigma^2}\right] \exp\left[\frac{i}{\hbar} p_0 x\right] \qquad \text{(B-1)}$$

Now, $a(p)$ is given by [see (Eq 2 2-9)]

$$\begin{aligned} a(p) &= \frac{1}{(2\pi\hbar)^{1/2}} \int_{-\infty}^{+\infty} \Psi(x, 0) \exp\left[-\frac{i}{\hbar} px\right] dx \\ &= \frac{1}{(2\pi\hbar)^{1/2}} \frac{2^{1/2}\sigma}{(\pi\sigma^2)^{1/4}} \left[\int_{-\infty}^{+\infty} e^{-\xi^2 + 2\alpha\xi}\, d\xi\right] \exp\left[-\frac{i}{\hbar}(p - p_0)x_0\right] \end{aligned} \qquad \text{(B-2)}$$

where

$$\xi = \frac{x - x_0}{2^{1/2}\sigma} \quad \text{and} \quad \alpha = -\frac{i}{2\hbar}(p - p_0) 2^{1/2}\sigma \qquad \text{(B-3)}$$

Thus

$$\begin{aligned} a(p) &= \frac{1}{(\pi\hbar)^{1/2}} \left(\frac{\sigma^2}{\pi}\right)^{1/4} e^{\alpha^2} \left[\int e^{-(\xi - \alpha)^2}\, d\xi\right] \exp\left[-\frac{i}{\hbar}(p - p_0)x_0\right] \\ &= \left(\frac{\sigma^2}{\pi\hbar^2}\right)^{1/4} \exp\left[-\frac{\sigma^2}{2\hbar^2}(p - p_0)^2 - \frac{i}{\hbar}(p - p_0)x_0\right] \end{aligned} \qquad \text{(B-4)}$$

Now, for a free particle, $E = p^2/2m$ and Eq (2 2-8) gives

$$\Psi(x, t) = \frac{1}{(2\pi\hbar)^{1/2}} \int_{-\infty}^{+\infty} a(p) \exp\left[\frac{i}{\hbar}\left(px - \frac{p^2}{2m} t\right)\right] dp \qquad \text{(B-5)}$$

Substituting the value of $a(p)$ from Eq (B-4) in Eq (B-5) we get

$$\Psi(x, t) = \left(\frac{\sigma^2}{\pi\hbar^2}\right)^{1/4} \frac{1}{(2\pi\hbar)^{1/2}} \int_{-\infty}^{+\infty} e^{-Ap^2 + Bp + C}\, dp \qquad \text{(B-6)}$$

where

$$A = \frac{\sigma^2}{2\hbar^2} + \frac{it}{2m\hbar} \tag{B-7}$$

$$B = \frac{\sigma^2}{\hbar^2} p_0 + \frac{i}{\hbar}(x - x_0) \tag{B-8}$$

$$C = -\frac{\sigma^2 p_0^2}{2\hbar^2} + \frac{i}{\hbar} p_0 x_0 \tag{B-9}$$

Now (see Eq A-26)

$$\int_{-\infty}^{\infty} e^{-Ap^2+Bp+C}\, dp = \left(\frac{\pi}{A}\right)^{1/2} \exp\left(\frac{B^2}{4A} + C\right) \tag{B-10}$$

Further,

$$\frac{B^2}{4A} = \frac{\sigma^2/2\hbar^2 - it/2m\hbar}{4[(\sigma^2/2\hbar^2)^2 + t^2/4m^2\hbar^2]}\left[\frac{\sigma^4}{\hbar^4} p_0^2 - \frac{(x - x_0)^2}{\hbar^2} + \frac{2i}{\hbar}(x - x_0)\frac{\sigma^2}{\hbar^2} p_0\right] \tag{B-11}$$

Thus

$$|\Psi(x, t)|^2 = \left(\frac{\sigma^2}{\pi\hbar^2}\right)^{1/2} \frac{1}{2\pi\hbar} \frac{\pi}{[(\sigma^2/2\hbar^2)^2 + t^2/4m^2\hbar^2]^{1/2}} \times \exp\left[\mathrm{Re}\left(\frac{B^2}{2A} + 2C\right)\right] \tag{B-12}$$

where Re[ ] represents the real part of the quantity inside the brackets On simplification, one immediately obtains

$$|\Psi(x, t)|^2 = \frac{1}{[\pi\sigma^2(1 + \hbar^2 t^2/m^2\sigma^4)]^{1/2}} \exp\left\{-\frac{[x - (x_0 + p_0 t/m)]^2}{\sigma^2(1 + \hbar^2 t^2/m^2\sigma^4)}\right\} \tag{B-13}$$

Equation (B-13) tells us that the envelope of the Gaussian wave packet moves with a velocity p_0/m, which represents the group velocity—see Eq (2 2-17) Further, its width increases by a factor $[1 + \hbar^2 t^2/m^2\sigma^4]^{1/2}$—see Fig 2 1 Thus, the envelope of a wave packet which is initially Gaussian remains Gaussian with the center of the packet moving with the group velocity p_0/m

C Planck's Law

In Section 6 2 we solved Maxwell's equations in a rectangular cavity and obtained the allowed frequencies of oscillation of the field in the

cavity [see Eq (6 2-16)] In Appendix D we will calculate the density of such modes and will show that the number of modes per unit volume in a frequency interval $d\omega$ will be given by [see Eq (D-10)]

$$p(\omega)\,d\omega = \frac{1}{\pi^2 c^3}\omega^2\,d\omega \tag{C-1}$$

In Section 8 2, we have shown that quantum mechanically, we can visualize the radiation field (inside a cavity) as consisting of an infinite number of simple harmonic oscillators (each oscillator corresponding to a particular mode of the cavity), and that the energy of each oscillator can take only the discrete values

$$E_{n_\lambda} = (n_\lambda + \tfrac{1}{2})\hbar\omega_\lambda, \qquad n_\lambda = 0, 1, 2, \tag{C-2}$$

[see Eq (8 2-59)] Now, according to Boltzmann's law, if E_{n_λ} is the energy of the λth mode in the nth excited state, then the probability p_{n_λ} of the system being in this state (at thermal equilibrium) is proportional to

$$\exp(-E_{n_\lambda}/k_B T)$$

where k_B is Boltzmann's constant and T is the absolute temperature Because the λth mode must be in one of the states, we must have

$$\sum_{n_\lambda=0}^{\infty} P_{n_\lambda} = 1 \tag{C-3}$$

To ensure this condition, we must have

$$P_{n_\lambda} = \frac{\exp(-E_{n_\lambda}/k_B T)}{\sum_{n_\lambda=0}^{\infty}\exp(-E_{n_\lambda}/k_B T)} \tag{C-4}$$

Substituting for E_{n_λ} from Eq (C-2), we have

$$P_{n_\lambda} = \frac{\exp(-n_\lambda x)}{\sum_{n_\lambda=0}^{\infty}\exp[-n_\lambda x]} \tag{C-5}$$

where

$$x = \frac{\hbar\omega_\lambda}{k_B T} \tag{C-6}$$

Carrying out the summation of the geometrical series in the denominator, we obtain

$$P_{n_\lambda} = [1 - e^{-x}]e^{-n_\lambda x} \tag{C-7}$$

The mean number of photons $\bar{n}_\lambda$ associated with the λth mode is

given by

$$\begin{aligned}\bar{n}_\lambda &= \sum_{n_\lambda=0}^{\infty} P_{n_\lambda} n_\lambda \\ &= (1 - e^{-x}) \sum_{n_\lambda=0}^{\infty} n_\lambda e^{-n_\lambda x} \\ &= -(1 - e^{-x}) \frac{d}{dx} \sum_{n_\lambda=0}^{\infty} e^{-n_\lambda x} \\ &= -(1 - e^{-x}) \frac{d}{dx} \left(\frac{1}{1 - e^{-x}} \right)\end{aligned}$$

or

$$\bar{n}_\lambda = \frac{1}{e^x - 1} = \frac{1}{e^{\hbar\omega_\lambda/k_B T} - 1} \tag{C-8}$$

Thus the mean energy associated with each mode will be

$$\bar{n}_\lambda \hbar\omega_\lambda = \frac{\hbar\omega_\lambda}{e^{\hbar\omega_\lambda/k_B T} - 1} \tag{C-9}$$

Consequently, if $u(\omega)$ represents the energy of the radiation field (per unit volume) in the frequency interval $d\omega$, then

$$\begin{aligned}u(\omega)\, d\omega &= (\text{energy associated with a mode}) \times (\text{number of modes per unit volume in the frequency interval } d\omega) \\ &= \frac{\hbar\omega}{e^{\hbar\omega/k_B T} - 1} \times \frac{1}{\pi^2 c^3} \omega^2\, d\omega\end{aligned}$$

or

$$u(\omega) = \frac{\hbar\omega^3}{\pi^2 c^3} \frac{1}{e^{\hbar\omega/k_B T} - 1} \tag{C-10}$$

which is the famous Planck's law For $k_B T \gg \hbar\omega$ one obtains

$$u(\omega) \simeq \frac{\omega^2}{\pi^2 c^3} k_B T \tag{C-11}$$

which is known as the Rayleigh–Jeans law On the other hand, at low temperatures, where $k_B T \ll \hbar\omega$ one obtains

$$u(\omega) \simeq \frac{\hbar\omega^3}{\pi^2 c^3} \exp\left(-\frac{\hbar\omega}{k_B T}\right) \tag{C-12}$$

which is known as Wien's law

D. The Density of States

In Section 6.2, we had solved Maxwell's equations in a rectangular cavity and had shown that the electric field inside the cavity is given by

$$E_x = E_{0x} \cos k_x x \sin k_y y \sin k_z z \tag{D-1}$$

with similar expressions for E_y and E_z [see Eq. (6.2-14)]. By imposing proper boundary conditions, the following allowed values of k_x, k_y, and k_z are obtained [see Eqs. (6.2-12) and (6.2-13)]

$$k_x = \frac{m\pi}{2a}, \qquad m = 0, 1, 2, \tag{D-2}$$

$$k_y = \frac{n\pi}{2b}, \qquad n = 0, 1, 2, \tag{D-3}$$

$$k_z = \frac{q\pi}{d}, \qquad q = 0, 1, 2, \tag{D-4}$$

Now, the number of modes whose x component of $\mathbf{k}$ lies between k_x and $k_x + dk_x$ would simply be the number of integers lying between $(2a/\pi)k_x$ and $(2a/\pi)(k_x + dk_x)$. This number would be approximately equal to $(2a/\pi)\, dk_x$. Similarly, the number of modes whose y and z components of $\mathbf{k}$ lie between k_y and $k_y + dk_y$ and k_z and $k_z + dk_z$ would, respectively, be

$$\frac{2b}{\pi}\, dk_y \quad \text{and} \quad \frac{d}{\pi}\, dk_z$$

Thus, there will be

$$\left(\frac{2a}{\pi}\, dk_x\right)\left(\frac{2b}{\pi}\, dk_y\right)\left(\frac{d}{\pi}\, dk_z\right) = \frac{V}{\pi^3}\, dk_x\, dk_y\, dk_z \tag{D-5}$$

modes in the range $dk_x\, dk_y\, dk_z$ of $\mathbf{k}$, here $V(= 2a \times 2b \times d)$ represents the volume of the cavity. Thus the number of modes per unit volume in the $\mathbf{k}$ space would be

$$\frac{V}{\pi^3}$$

If $P(k)\, dk$ represents the number of modes whose $|\mathbf{k}|$ lies between k and $k + dk$, then

$$P(k)\, dk = 2 \times \frac{1}{8} \times \frac{V}{\pi^3} 4\pi k^2\, dk \tag{D-6}$$

where the factor $4\pi k^2\, dk$ represents the volume element (in the $\mathbf{k}$ space) lying between k and $k + dk$ and the factor 1/8 is due to the fact that k_x,

k_y, and k_z can take only positive values [see Eqs (D-2)–(D-4)] so that while counting the modes in the **k** space we must consider only the positive octant, the factor of 2 corresponds to the fact that corresponding to a particular value of **k**, there are two independent modes of polarization If $p(k)\,dk$ represents the corresponding number of modes per unit volume then

$$p(k)\,dk = \frac{1}{\pi^2} k^2\,dk \tag{D-7}$$

Now, if $p(\nu)\,d\nu$ represents the number of modes (per unit volume) in the frequency interval $d\nu$, then

$$p(\nu)\,d\nu = p(k)\,dk = \frac{1}{\pi^2} k^2\,dk = \frac{1}{\pi^2}\left(\frac{2\pi\nu}{c}\right)^2 \frac{2\pi}{c}\,d\nu \tag{D-8}$$

or

$$p(\nu)\,d\nu = \frac{8\pi\nu^2}{c^3}\,d\nu \tag{D-9}$$

which is identical to Eq (6 2-17) In deriving Eq (D-9) we have used the relation $k = 2\pi\nu/c$ Transforming to the ω space, we obtain

$$p(\omega)\,d\omega = \frac{1}{\pi^2 c^3}\,\omega^2\,d\omega \tag{D-10}$$

It should be mentioned that in the derivation of the above formula we have assumed $|\mathbf{k}|$ to lie between k and $k + dk$ and have integrated over all the directions of **k** If, however, we are interested in the number of modes for which **k** lies between k and $k + dk$ but the direction of **k** lies in the solid angle $d\Omega$, then the number of such modes would be given by

$$N(k)\,dk\,d\Omega = \frac{1}{8} \times \frac{V}{\pi^3} \times 4\pi k^2\,dk \times \frac{d\Omega}{4\pi} \tag{D-11}$$

and we have not taken into account the two independent states of polarization Thus,

$$N(\omega)\,d\omega\,d\Omega = N(k)\,dk\,d\Omega = \frac{V}{8\pi^3} k^2\,dk\,d\Omega$$

or

$$N(\omega)\,d\omega\,d\Omega = \frac{V\omega^2}{8\pi^3 c^3}\,d\omega\,d\Omega \tag{D-12}$$

which is identical to Eq (8 5-30)

E. Fourier Transforming Property of a Lens

In this Appendix, we will show that the amplitude distribution at the back focal plane of a lens is the spatial Fourier transform of the object field distribution at the front focal plane In order to show this, we must first determine the effect of a lens on an incident field

Consider an object point O which is at a distance d_1 from an aberrationless thin lens of focal length f (see Fig E 1) We know that under geometrical optics approximation, the lens images the point O at the point I which is at a distance d_2 from the lens where

$$\frac{1}{d_2} = \frac{1}{f} - \frac{1}{d_1} \tag{E-1}$$

Thus, the lens transforms an incident diverging spherical wave emanating from O into a converging spherical wave converging to the point I The phase of the incident diverging spherical wave from O can be written as

$$\begin{aligned} \exp(-ikr) &= \exp[-ik(x^2 + y^2 + d_1^2)^{1/2}] \\ &\simeq \exp(-ikd_1)\exp\left(-ik\frac{x^2 + y^2}{2d_1}\right) \end{aligned} \tag{E-2}$$

where we have assumed a time dependence of the form $e^{i\omega t}$ so that $\exp[i(\omega t - kr)]$ and $\exp[i(\omega t + kr)]$ represent, respectively, a diverging and a converging spherical wave and in writing the last expression, we have assumed that $x, y \ll d_1$ The wave after passing through the lens becomes a converging spherical wave given by

$$\exp(+ikd_2)\exp\left(+ik\frac{x^2 + y^2}{2d_2}\right) \tag{E-3}$$

where we have assumed $x, y \ll d_2$ and the positive sign refers to the fact that the wave is a converging spherical wave Hence, if we represent by p_L

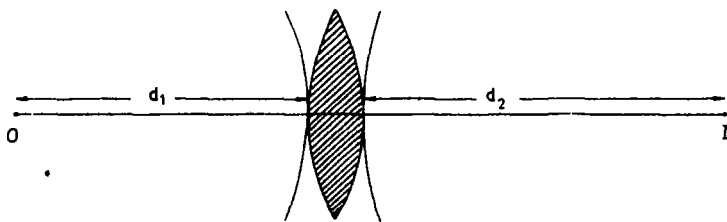



Fig E 1 A spherical wave emanating from an object point which is at a distance d_1 from a lens is converted into a converging spherical wave converging towards the point I at a distance d_2 satisfying Eq (E-1)

the effect of the lens on the incident field distribution, then

$$\exp(-ikd_1)\exp\left[-\frac{ik}{2d_1}(x^2+y^2)\right]p_L = \exp(+ikd_2)\exp\left[+\frac{ik}{2d_2}(x^2+y^2)\right]$$

or

$$p_L = \exp[ik(d_2+d_1)]\exp\left[\frac{ik}{2}\left(\frac{1}{d_1}+\frac{1}{d_2}\right)(x^2+y^2)\right] \qquad \text{(E-4)}$$

Using Eq (E-1) and neglecting the first factor on the right-hand side of Eq (E-4) as it is a constant phase factor, one obtains

$$p_L = \exp\left[+\frac{ik}{2f}(x^2+y^2)\right] \qquad \text{(E-5)}$$

Thus, the effect of a thin lens of focal length f is to multiply the incident field distribution by the factor p_L given by Eq (E-5)

In order to obtain the field distribution at the back focal plane, we would also require the effect of propagation through space, this is indeed given by Eq (6 8-4), namely,

$$g(x, y, z) \approx \frac{i}{\lambda z} e^{-ikz} \iint f(x', y') \exp\left\{-\frac{ik}{2z}[(x-x')^2 + (y-y')^2]\right\} dx'\, dy' \qquad \text{(E-6)}$$

where $f(x', y')$ is the field distribution on the plane $z = 0$ and $g(x, y, z)$ is the field at the point x, y, z

Let us now consider a field distribution $f(x, y)$ on the front focal plane (P_1) of a lens of focal length f (see Fig E 2) The field on the plane P_3 would be given by Eq (E-6) with z replaced by f, i e ,

$$\frac{i}{\lambda f} e^{-ikf} \iint f(\xi, \eta) \exp\left\{-\frac{ik}{2f}[(x'-\xi)^2 + (y'-\eta)^2]\right\} d\xi\, d\eta$$

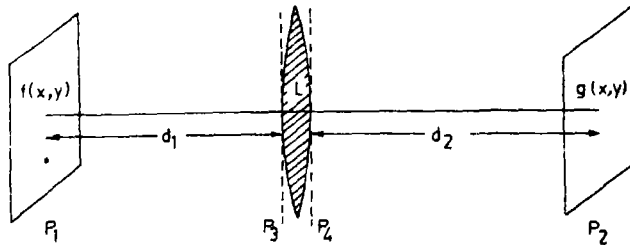



Fig E 2 The field distribution on the back focal plane (P_2) of a lens of focal length f is the Fourier transform of the field distribution on the front focal plane P_1

This field on passing through the lens becomes

$$\frac{i}{\lambda f} e^{-ikf} \iint f(\xi, \eta) \exp\left\{-\frac{ik}{2f}[(x'-\xi)^2 + (y'-\eta)^2]\right\} d\xi\, d\eta \times \exp\left[\frac{ik}{2f}(x'^2 + y'^2)\right]$$

Thus, the field on the back focal plane P_2 would be

$$g(x, y) = \left(\frac{i}{\lambda f}\right)^2 e^{-2ikf} \iiiint f(\xi, \eta) \exp\left\{-\frac{ik}{2f}[(x'-\xi)^2 + (y'-\eta)^2]\right\} \times \exp\left\{-\frac{ik}{2f}[(x-x')^2 + (y-y')^2]\right\} \times \exp\left[\frac{ik}{2f}(x'^2 + y'^2)\right] d\xi\, d\eta\, dx'\, dy' \tag{E-7}$$

which on simplification gives

$$\begin{aligned} g(x, y) &= -\frac{1}{\lambda^2 f^2} e^{-2ikf} \exp\left[-\frac{ik}{2f}(x^2 + y^2)\right] \iint d\xi\, d\eta\, f(\xi, \eta) \\ &\quad \times \exp\left[-\frac{ik}{2f}(\xi^2 + \eta^2)\right] \int \exp\left\{-\frac{ik}{2f}[x'^2 - 2x'(x+\xi)]\right\} dx' \\ &\quad \times \int \exp\left\{-\frac{ik}{2f}[y'^2 - 2y'(y+\eta)]\right\} dy' \\ &= -\frac{1}{\lambda^2 f^2} \frac{\pi 2 f \lambda}{2\pi i} e^{-2ikf} \exp\left[-\frac{ik}{2f}(x^2 + y^2)\right] \iint d\xi\, d\eta\, f(\xi, \eta) \\ &\quad \times \exp\left[-\frac{ik}{2f}(\xi^2 + \eta^2)\right] \\ &\quad \times \exp\left[+\frac{ik}{2f}(x^2 + \xi^2 + 2x\xi + y^2 + \eta^2 + 2y\eta)\right] \\ &= \frac{i}{\lambda f} e^{-2ikf} \iint d\xi\, d\eta\, f(\xi, \eta) \exp\left[2\pi i\left(\frac{x\xi}{\lambda f} + \frac{y\eta}{\lambda f}\right)\right] \end{aligned} \tag{E-8}$$

where we have used the following result

$$\int_{-\infty}^{\infty} e^{-px^2+qx}\, dx = \left(\frac{\pi}{p}\right)^{1/2} \exp\left(\frac{q^2}{4p}\right) \tag{E-9}$$

Thus, apart from a constant phase factor, the amplitude distribution at the back focal plane of a lens is nothing but the Fourier transform of the amplitude distribution at the front focal plane evaluated at spatial frequencies $x/\lambda f$ and $y/\lambda f$

F. The Natural Lineshape Function

A spontaneous transition from a state b to a lower energy state a does not give radiation at a single frequency ω_{ba}. A finite lifetime τ of the excited state gives it an energy width of the order $\hbar/\tau$ so that the emitted radiation has a frequency distribution. This argument may be made more precise by considering the simple case of an atom with only two states a and b ($E_b > E_a$) undergoing spontaneous transition from b to a. In this Section we will calculate the frequency distribution of the emitted radiation from such a spontaneous transition, the analysis is based on the treatment given by Heitler (1954), Section 18 and is known as Weisskopf–Wigner theory of the natural line width.

It is assumed that at $t = 0$, the atom is in the excited state b and the radiation field has no photons. We denote this state by

$$|1\rangle = |b, 0, 0, 0, \ldots\rangle \tag{F-1}$$

[compare Eqs. (8.5-15) and (8.5-16)]. The atom makes a transition to the state a emitting a photon of frequency ω_λ in the mode characterized by λ [see Eq. (8.2-61)]. We denote this state by

$$|a\lambda\rangle = |a, 0, 0, \ldots 1_\lambda, 0, \ldots\rangle \tag{F-2}$$

If we denote the corresponding probability amplitudes by

$$C_1(t) \quad \text{and} \quad C_{a\lambda}(t)$$

then

$$\begin{aligned} C_1(0) &= 1 \\ C_{a\lambda}(0) &= 0 \end{aligned} \tag{F-3}$$

Further [see Eq. (8.5-11)]

$$\begin{aligned} i\hbar \frac{dC_1}{dt} &= \sum_\lambda \langle 1| H' |a\lambda\rangle e^{i(W_1 - W_{a\lambda})t/\hbar} C_{a\lambda}(t) \\ &= \sum_\lambda \langle 1| H' |a\lambda\rangle e^{i(\omega_0 - \omega_\lambda)t} C_{a\lambda}(t) \end{aligned} \tag{F-4}$$

and

$$i\hbar \frac{dC_{a\lambda}}{dt} = \langle a\lambda| H' |1\rangle e^{-i(\omega_0 - \omega_\lambda)t} C_1(t) \tag{F-5}$$

where

$$\hbar\omega_0 = E_b - E_a \tag{F-6}$$

and

$$W_1 - W_{a\lambda} = E_b - (E_a + \hbar\omega_\lambda) = \hbar(\omega_0 - \omega_\lambda) \tag{F-7}$$

We try to solve Eqs (F-4) and (F-5) by assuming†

$$C_1(t) = e^{-\gamma t/2} \tag{F-8}$$

which satisfies the condition $C_1(0) = 1$ It may be noted that since $|C_1(t)|^2$ represents the probability of finding the atom in the upper state, the quantity $1/\gamma$ represents the mean lifetime of the state If we substitute for $C_1(t)$ in Eq (F-5) and carry out the integration we get

$$C_{a\lambda}(t) = \frac{\langle a\lambda|\,H'\,|1\rangle}{\hbar}\,\frac{\exp\{-[i(\omega_0 - \omega_\lambda) + \gamma/2]t\} - 1}{(\omega_0 - \omega_\lambda) - i\gamma/2} \tag{F-9}$$

We substitute Eqs (F-8) and (F-9) in (F-4) to obtain

$$-i\hbar\frac{\gamma}{2} = \sum_\lambda \frac{|\langle H'\rangle|^2}{\hbar}\,\frac{1-\exp[i(\omega_0 - \omega_\lambda)t + \gamma t/2]}{(\omega_0 - \omega_\lambda) - i\gamma/2} \tag{F-10}$$

where $\langle H'\rangle \equiv \langle a\lambda|\,H'\,|1\rangle$ The summation on the right-hand side of the above equation is over a number of states λ with very nearly the same frequency Under these circumstances, the summation can be replaced by an integral,

$$\sum_\lambda \rightarrow \iint N(\omega)\,d\omega\,d\Omega \tag{F-11}$$

where $N(\omega)$ represents the density of states [see Eq (8 5-30)] Equation (F-10) then becomes

$$\gamma = \frac{2i}{\hbar}\left[\frac{1}{\hbar}\iint |\langle H'\rangle|^2\,N(\omega)J(\omega)\,d\omega\,d\Omega\right] \tag{F-12}$$

where

$$J(\omega) = \frac{1-\exp[i(\omega_0 - \omega)t]e^{\gamma t/2}}{(\omega_0 - \omega) - i\gamma/2} \tag{F-13}$$

We will assume that $\gamma \ll \omega_0$, i e , the inverse of the lifetime ($\leqslant 10^9\,\text{sec}^{-1}$) is much smaller than the characteristic frequencies ($\sim 10^{15}\,\text{sec}^{-1}$), this will indeed follow from the final result Under this assumption γ can be

† It should be mentioned that the present treatment is consistent with the treatment given in Section 8 5 Indeed, in Section 8 5 we have shown that A [as given by Eq (8 5-36)] represents the probability per unit time for the spontaneous emission to occur, the decay given by Eq (F-8) follows with $\gamma = A$, this is explicitly shown later in this Section

neglected in Eq (F-13) to obtain

$$J \simeq \frac{1 - e^{i(\omega_0 - \omega)t}}{\omega_0 - \omega}$$

$$\simeq \frac{1 - \cos[(\omega_0 - \omega)t]}{(\omega_0 - \omega)} - i\frac{\sin[(\omega_0 - \omega)t]}{(\omega_0 - \omega)} \tag{F-14}$$

The first term has very rapid variation around $\omega = \omega_0$ and gives negligible contribution to any integral over ω except around $\omega \approx \omega_0$, where the function itself vanishes Thus this term will lead to a small imaginary value of γ We are in any case interested in the real part of γ (which would give the lifetime, etc) which will be given by

$$\gamma \simeq \frac{2}{\hbar}\left[\frac{1}{\hbar}\int |\langle H'\rangle|^2 N(\omega_0)\, d\Omega \int_{-\infty}^{\infty} \frac{\sin(\omega_0 - \omega)t}{(\omega_0 - \omega)}\, d\omega\right] \tag{F-15}$$

We have made two approximations here First, we have pulled out factors which are essentially constant in the neighborhood of $\omega \approx \omega_0$ Second, the limits of the integral have been taken from $-\infty$ to $+\infty$ since in any case the contribution vanishes except around ω_0 Thus

$$\gamma \simeq \frac{2\pi}{\hbar}\left[\frac{1}{\hbar}\int |\langle H'\rangle|^2 N(\omega_0)\, d\Omega\right]$$

$$= \frac{2\pi}{\hbar}\frac{1}{\hbar}\frac{V\omega_0^2}{8\pi^3 c^3}\int |\langle H'\rangle|^2\, d\Omega \tag{F-16}$$

Now

$$\langle H'\rangle = \langle a\lambda|\, H'\,|1\rangle$$

$$= \langle a, 0, 0, \quad 1_\lambda, 0, \quad |(-ie)$$

$$\times\sum_\lambda \left(\frac{\hbar\omega_\lambda}{2\varepsilon_0 V}\right)^{1/2}[a_\lambda e^{i\mathbf{k}_\lambda\cdot\mathbf{r}} - \bar{a}_\lambda e^{-i\mathbf{k}_\lambda\cdot\mathbf{r}}]\hat{\mathbf{e}}_\lambda\cdot\mathbf{r}\,|b, 0, 0, \quad, 0_\lambda, \quad\rangle$$

$$= (ie)\left(\frac{\hbar\omega_\lambda}{2\varepsilon_0 V}\right)^{1/2}\langle a|\, e^{-i\mathbf{k}_\lambda\cdot\mathbf{r}}\mathbf{r}\,|b\rangle\cdot\hat{\mathbf{e}}_\lambda$$

$$\simeq ie\left(\frac{\hbar\omega_\lambda}{2\varepsilon_0 V}\right)^{1/2}\langle a|\,\mathbf{r}\,|b\rangle\cdot\hat{\mathbf{e}}_\lambda \tag{F-17}$$

where in the last step we have used the dipole approximation [see th discussion after Eq (8 5-20)] On substitution in Eq (F-16), we get

$$\gamma \simeq \frac{1}{2\pi}\left[\frac{e^2}{4\pi\varepsilon_0\hbar c}\right]\frac{\omega_0^3}{c^2}\int|\hat{\mathbf{e}}_\lambda\cdot\langle a|\mathbf{r}|b\rangle|^2\,d\Omega \qquad \text{(F-18)}$$

[compare Eq (8 5-34)]. On carrying out the integration and summin over the two states of polarization we get [see Eq (8 5-36)]

$$\gamma = A \qquad \text{(F-19)}$$

i e , γ is simply the Einstein A coefficient corresponding to spontaneou emissions as it indeed should be!

Now the probability of a photon being emitted in the mode λ is give by [see Eq (F-9)]

$$|C_{a\lambda}(t=\infty)|^2 = \frac{|\langle H'\rangle|^2}{\hbar^2}\frac{1}{(\omega_0-\omega_\lambda)^2+\gamma^2/4} \qquad \text{(F-20)}$$

If we multiply the above equation by $N(\omega)\,d\omega\,d\Omega$, which gives th number of modes in the frequency interval ω to $\omega + d\omega$ and in the soli angle $d\Omega$, and carry out the integration over $d\Omega$ and sum over the tw states of polarization, we would get

$$g(\omega)\,d\omega = \frac{\gamma}{2\pi}\frac{d\omega}{(\omega_0-\omega)^2+\gamma^2/4} \qquad \text{(F-2}$$

where we have used Eq (F-18) The above equation indeed gives us th probability that the spontaneously emitted photon has its frequenc between ω and $\omega + d\omega$, which is nothing but the Lorentzian line shap [compare Eq (3 5-19)] Notice that the above expression is normalize which implies that the probability that the atom makes a spontaneou transition from the upper state to the lower state is unity (as $t \to \infty$), as indeed should be

References

Alfano R R and Shapiro, S L (1975) Ultrashort phenomena, *Phys Today* **31**(7), 30

Arditty, H , Papuchon, M , Puech, C , and Thyagarajan K (1980), Recent developments in guided wave optical rotation sensors, presented at the Conference on Integrated and Guided-Wave Optics, Nevada, January 28–30

Arecchi, F T (1976), Quantum optics and photon statistics in *Lasers and Their Applications* (A Sona, ed), p 497, Gordon and Breach, New York

Arecchi, F T., Berne, A , and Burlamacchi P (1966a) High-order fluctuations in a single mode laser field, *Phys Rev Lett* **16**, 32

Arecchi, F T , Gatti, E , and Sona A (1966b) Time distribution of photons from coherent and Gaussian sources, *Phys Lett* **20**, 27

Baldwin, G C (1969) *An Introduction to Nonlinear Optics*, Plenum Press, New York

Barnoski, M (1976) *Fundamentals of Optical Fiber Communications*, Academic Press, New York

Baym, G (1969), *Lectures on Quantum Mechanics*, W A Benjamin, Inc , New York

Bennett, W R (1962), Hole burning effects in a He–Ne optical Maser, *Phys Rev* **126**, 580

Bernhardt, A F , Duerre, D E , Simpson, J R , and Wood, L L (1974), Separation of isotopes by laser deflection of atomic beam, I Barium, *Appl Phys Lett* **25**, 617

Bleuler, E , and Goldsmith, G J (1952), *Experimental Nucleonics*, Holt, Rinehart and Winston, New York

Booth, L A , Freiwald, D A , Frank, T G , and Finch, F T (1976), Prospects of generating power with laser driven fusion, *Proc IEEE* **64**, 1460

Born, M , and Wolf, E (1975), *Principles of Optics*, Pergamon Press, Oxford

Boyd, G D , and Gordon, J P (1961), Confocal multimode resonator for millimeter through optical wavelength masers, *Bell Syst Tech J* **40**, 489

Brouwer W (1964), *Matrix Methods in Optical Instrumental Design*, W A Benjamin Inc , New York

Brueckner, K A , and Jorna, S (1974), Laser driven fusion, *Rev Mod Phys* **46**(2), 325

Busch, G E , and Rentzepis, P M (1976), Picosecond chemistry, *Science* **194**, 276

Cassent, D (ed) (1978), *Optical Data Processing*, Springer-Verlag, Berlin

Cattermole, K W (1969), *Principles of Pulse Code Modulation*, Iliffe Books Ltd London

Charschan, S S (1972), *Lasers in Industry*, Van Nostrand Reinhold Co , New York

Chynoweth, A. G (1976), Lightwave communications The fiber lightguide *Phys Today* **29**(5), 28

Collier, R. J , Burckhardt, C B , and Lin, L. H. (1971), *Optical Holography*, Academic Press, New York

Condon, E U , and Shortley, G H (1935), *Theory of Atomic Spectra*, Cambridge Univ Press, Cambridge

Dirac, P A M (1958a), *The Principles of Quantum Mechanics*, Oxford Univ Press, London

Dirac, P A. M (1958b), *Quantum theory of emission and absorption in quantum electrodynamics* (J Schwinger, ed), Dover Publications, New York

Einstein, A (1917), On the quantum theory of radiation, *Phys. Z* **18**, 121 (This paper has been reprinted in *Laser Theory* by F S Barnes, IEEE Press, New York, 1972.)

Fox, A G , and Li, T (1961), Resonant modes in a maser interferometer, *Bell. Syst. Tech J* **40**, 453

Franken, P A , Hill, A E , Peters, C W , and Weinreich, G (1961), Generation of optical harmonics, *Phys Rev Lett.* **7**, 118

Gagliano, F P , Lumley, R M , and Watkins, L S (1969), Lasers in industry, *Proc IEEE* **57**, 114

Gallawa, R (1979), Optical systems a review I U S and Canada initial results reported, *IEEE Spectrum* **16**(10), 71

Gerrard, A , and Burch, J M (1975), *An Introduction to Matrix Methods in Optics*, Wiley, New York

Ghatak, A K , and Lokanathan, S (1977), *Quantum Mechanics*, Macmillan, New Delhi

Ghatak, A K , and Thyagarajan, K (1978), *Contemporary Optics*, Plenum Press, New York

Ghatak, A , and Thyagarajan, K (1980), Graded index optical waveguides A review, *Progress in Optics* (E Wolf, ed), Vol XVIII, North-Holland Publ Co , Amsterdam

Gloge, D (1976), *Optical Fiber Technology*, IEEE Press, New York

Gloge, D (1979), Inside the new region second generation fiber systems, *Opt Spectra* **131**(11), 50

Goldman, L , and Rockwell, R J (1971), *Lasers in Medicine*, Gordon and Breach Science Publishers Inc , New York

Goldstein, H (1950), *Classical Mechanics*, Addison-Wesley, Reading, Massachusetts

Goldstein, R , and Goss, W G (1979), Fiber optic rotation sensor (FORS), Laboratory performance evaluation, *Opt. Eng* (July–Aug) **18**(4), 381

Gopal, E S R (1974) *Statistical Mechanics and Properties of Matter*, Wiley, New York

Gordon, J P , Zeiger, H J , and Townes, C H (1955), The Maser—New type of microwave amplifier, frequency standard and spectrometer, *Phys Rev* **99**, 1264

Hall, F F (1974), Laser systems for monitoring the environment, in *Laser Applications*, Vol 2 (M Ross, ed), Academic Press, New York

Harry, J E (1974), *Industrial Lasers and Their Applications*, McGraw Hill, London

Heitler, W (1954), *Quantum Theory of Radiation*, Oxford University Press, London

Jacobs, S F (1979), How monochromatic is laser light?, *Am J Phys* **47**, 597

Jacobs, I , and Miller, S E (1977), Optical transmission of voice and data, *IEEE Spectrum* **14**(2), 32

Jenkins, F A , and White, H E (1957), *Fundamentals of Optics*, McGraw-Hill, New York

Khokhlov, R V , Akhmanov, S A , and Sukhorukov, A P (1976), Self-focusing, self-defocusing and self-modulation of light beams in nonlinear media, in *Laser Handbook* (F T Arecchi and E O Schulz-Dubois, eds), North-Holland Publ Co , Amsterdam

Kidder, R E (1973), Some aspects of controlled fusion by use of lasers, in *Fundamental and Applied Laser Physics* (M. S Feld, A Javan, and N A. Kurnit, eds), John Wiley, New York

Kogelnik, H , and Li, T (1966), Laser beams and resonators, *Appl. Opt.* **5**, 1550

Kressel, H. (1979), Semiconductor diode lasers, in *Quantum Electronics*, Part A, Vol 15 of *Methods of Experimental Physics* (C L. Tang, ed), Academic Press, New York

Kressel, H., and Butler, J K. (1977), *Semiconductor Lasers Heterojunction LEDs*, Academic Press, New York.

Kressel, H , Ladany, I , Ettenberg, M., and Lockwood, H (1976), Lightwave communications Light sources, *Phys. Today* **29**(5), 38

Lamb, W. E. (1964), Theory of an optical maser, *Phys. Rev* **134**, A1429.

Lehr, C G (1974), Laser tracking systems, in *Laser Applications*, Vol 2 (M. Ross, ed), Academic Press, New York

Letokhov, V S (1977), Lasers in research Photophysics and photochemistry, *Phys. Today* **30**(5), 23

Lin, C , Stolen, R H , and Cohen, L G (1977a), A tunable 1 1 μm fiber Raman oscillator, *Appl Phys. Lett.* **2**, 97

Lin, C , Stolen, R H , French, W G , and Malone, T G (1977b), A CW tunable near infrared (1 085–1 175 μm) Raman oscillator, *Opt Lett* **1**, 96

Loudon, R (1973), *The Quantum Theory of Light*, Clarendon Press, Oxford

McFarlane, R A , Bennet, W R , and Lamb W E (1963), Single mode tuning dip in the power output of an He–Ne optical maser, *Appl. Phys Lett* **2**, 189

Maitland, A. and Dunn, M H (1969), *Laser Physics*, North-Holland Publ Co , Amsterdam

Mandel, L (1958), Fluctuations of photon beams and their correlations, *Proc Phys Soc* **72**, 1037

Mandel, L (1959), Fluctuations of photon beams The distribution of the photo-electrons, *Proc Phys Soc* **74**, 233

Mandel, L , and Wolf, E (1970) *Selected Papers on Coherence and Fluctuations of Light*, Vols I and II, Dover Publications, New York

Mayer, S W , Kwok, M A , Gross, R W F and Spencer, D J (1970), Isotope separation with the CW hydrogen fluoride laser, *Appl Phys Lett* **17**, 516

Mehta, C L (1970), Theory of photoelectron counting, in *Progress in Optics* (E. Wolf, ed), Vol VIII, p 375, North-Holland Publ Co , Amsterdam

Melchoir, H (1977), Detectors for light wave communication, *Phys Today* **30**(11), 32.

Midwinter, J E (1979a), *Optical Fibers for Transmission*, John Wiley, New York

Midwinter, J E (1979b), Optical systems a review III, Europe business benefits, *IEEE Spectrum* **16**(10), 76

Minck, R W , Terhune, R W , and Wang, C C (1966), Nonlinear optics, *Appl. Opt* **5**, 1595

Miya, T , Terunuma, Y , Hosaka, T , and Miyashita, T (1979), Ultra low loss single-mode fibers at 1 55 μm, *Rev Electr Commun. Lab* **27**(7–8), 497

Nelson, D F , and Collins, R. J (1961), Spatial coherence in the output of an optical maser, *J Appl Phys* **32**, 739

Nuckolls, J , Wood, L., Thiessen, A , and Zimmerman, G (1972), Laser compression of matter to super high densities Thermonuclear applications, *Nature* **239**, 139

Nuese, C J , Olsen, G H , Ettenberg, M , Gannon, J J , and Zamarowsky, T J (1976), CW room temperature $In_xGa_{1-x}As/In_yGa_{1-y}P$ 1 06 μm laser, *Appl. Phys Lett.* **29**, 807

Nussbaum, A. (1968), *Geometric Optics An Introduction*, Addison-Wesley, Reading, Massachusetts

Pal, B P (1979), Optical communication fiber waveguide fabrication A review, *Fiber Int. Opt.* **2**(2), 195

Patel, C K N (1966), Optical harmonic generation in the infrared using a CO_2 laser, *Phys. Rev Lett* **16**, 613

Patel, C K N , Slusher, R E , and Flurry, P A. (1966), Optical nonlinearities due to mobile carriers in semiconductors, *Phys. Rev. Lett.* **17**, 1011

Personick, S D (1976), Photodetectors for fiber systems, in *Fundamentals of Optical Fiber Communications* (M K Barnoski, ed), p 155, Academic Press, New York

Post, E J (1967), Sagnac effect, *Rev Mod. Phys.* **39**, 475

Post, R F (1973), Prospects for fusion power, *Phys. Today* **26**(4), 30

Powell, J L, and Craseman, B (1961), *Quantum Mechanics*, Addison-Wesley, Reading, Massachusetts

Ribe, F L (1975), Fusion reactor systems, *Rev Mod Phys.* **47**, 7

Salzman, W R. (1971), Time evolution of simple quantum mechanical systems II Two state system in intense fields, *Phys Rev Lett* **26**, 220

Schafer, F P (1973), *Dye Lasers*, Springer Verlag, Berlin

Schawlow, A L., and Townes, C H (1958) Infrared and optical masers, *Phys Rev* **112**, 1940

Shapiro, S L (1977), *Ultrashort Light Pulses Picosecond Techniques and Applications*, Springer Verlag, Berlin

Shimada, S (1979), Optical systems a review, II Japan unusual applications, *IEEE Spectrum* **16**(10), 74

Slepian, D, and Pollack, H O (1961), Prolate spheroidal wave functions—Fourier analysis and uncertainty—I, *Bell Syst Tech J* **40**, 43

Smith, P W (1972), Mode selection in lasers, *Proc IEEE* **60**, 422

Sodha, M. S, and Ghatak, A K (1977), *Inhomogeneous Optical Waveguides*, Plenum Press, New York

Sodha, M. S., Ghatak, A K, and Tripathi, V K (1974), *Self-Focusing of Laser Beams in Dielectrics, Plasmas and Semiconductors*, Tata McGraw Hill New Delhi

Sodha, M S, Ghatak, A K, and Tripathi, V K (1976), Self-focusing of laser beams in plasmas and semiconductors, *Progress in Optics* Vol XIII (E Wolf, ed), North-Holland Publ Co, Amsterdam

Steensma, P D, and Mondrick, A (1979), Battlefield fiber network, *Laser Focus* **15**(7), 52

Stickley, C M (1978), Laser fusion, *Phys Today* **31**(5), 50

Stolen, R H, Ippen, E P, and Tynes, A R (1972), Raman oscillation in glass optical waveguides, *Appl Phys. Lett.* **20**, 62

Stolen, R H, Lin, C, and Jain, R K (1977), A time dispersion tuned fiber Raman oscillator, *Appl. Phys Lett* **30**, 340

Susskind, L, and Glogower, J (1964), Quantum mechanical phase and time operator, *Physics* **1**, 49

Svelto, O (1975), Self focusing, self trapping and self modulation of laser beams, *Progress in Optics*, Vol. XII (E Wolf, ed), North-Holland Publ Co, Amsterdam

Tait, J H (1964), *Neutron Transport Theory*, Longmans, London

Tsujiuchi, J, Matsuda, K, and Takeja, N (1971), Correlation techniques by holography and its application to fingerprint identification, in *Applications of Holography* (E. S Barrekette, W E Kock, T Ose, J Tsujiuchi, and G W Stroke, eds), p 247, Plenum Press, New York

Vali, V, and Shorthill, R W (1977), Ring interferometer 950 m long, *Appl Opt* **16**, 290

Weaver, L. A (1971), Machining and welding applications, in *Laser Applications* (M Ross, ed), Vol I, p 201, Academic Press, New York

Zare, R. N (1977), Laser separation of isotopes, *Sci. Am* **236**(2), 86

Zernike, F, and Midwinter, J E (1973), *Applied Nonlinear Optics*, John Wiley and Sons, New York

Zuber, K (1935), *Nature* **136**, 796

Index

NICE Book FOR "LASER"

VIKAS